10-1-1981

A New Approach to Small Sample Theory

Peter C.B. Phillips

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 608

A NEW APPROACH TO SMALL SAMPLE THEORY

Peter C. B. Phillips

October 9, 1981

1. INTRODUCTION

Analytical research on the small sample properties of econometric methods of estimation and testing has taken three main directions. The first of these has involved the mathematical task of extracting the form of the probability density function (pdf) or distribution function (df) of the relevant statistic under stated assumptions concerning the structure of the model and the stochastic properties of the errors driving its equations. The second has been concerned with characterizing the distribution by the analysis and approximation of its moments. This has included work on the question of the existence of moments and conditions for asymptotic approximations to them to be valid. Finally, there has in recent years been a growing literature concerned with the derivation of direct approximations to the distributions themselves. These approximations have frequently been obtained by truncating asymptotic series expansions after a small number of terms. This literature has also tackled problems such as the validity of the expansions as asymptotic series and the actual numerical performance of the approximations in a variety of situations. In addition and largely in parallel to this analytical research, are the experimental investigations, which have continued traditions established in the 1950's and 1960's with an attempt to improve certain features of the design and efficiency of the experiments, together with the means by which the results of the experiments are characterized.

In contrast to this research, the foundation for estimation and inferential procedures in practical econometric work rests almost exclusively on asymptotic theory, particularly in the areas of simultaneous

equations, non linear regressions, time series methods, limited dependent variable and qualitative response models. This foundation has been strengthened by the conventional coding of asymptotic statistics in computer regression packages. The presence of asymptotic statistics in the printouts of these packages and the typical absence of any finite sample alternative has no doubt entrenched the practice of a sole reliance on asymptotic theory in most empirical econometric work.

The ultimate objective of the research in small sample theory has been to relieve the empirical worker from this heavy reliance on asymptotic theory; but there has, as yet, been no substantial payoff to this research in terms of applied econometric practice. This situation is most likely to change in important ways during the 1980's. In part, this is because the rather specialized results of the early research have recently given way to general theories and a powerful technical machinery which will make it easier to transmit results and methods to the applied econometrician in the precise setting of the model and the data set with which he is working. And, in part, this is because improvements in computing now make it feasible to incorporate into existing regression software subroutines which will provide the essential vehicle for this transmission. An integral part of this process will be played by direct approximations to the sampling or posterior distributions of interest in an applied study. Of the various directions to existing research in small sample theory, it is this approach which, in my view, offers most promise in terms of results that can be used directly in applied econometric work. Not only in sampling theoretic, but also in Bayesian problems.

Direct approximations to the distributions of interest in an applied study can be utilized in the decisions that have to be made concerning

a choice of estimator and the specification of a critical region in a statistical test; they can also be used to reduce and characterize the multi-dimensional posterior distributions that arise in Bayesian inference. To illustrate the first application we can take the case of the Edgeworth approximation where the theory that has been developed is sufficiently general for the formulae to be incorporated in general purpose regression packages at the estimation stage itself. This was envisaged by Sargan [42]. The formulae can be algebraically built in to a subroutine which is called after the model is read in and perhaps after some estimation is completed. The subroutine would return tabulations and graphical plots of the approximate distributions of the estimators and, in the case of test statistics, tail area probabilities corresponding to given percentage points based on the known asymptotic distribution or an approximate critical region corresponding to a stated test size. This information would be provided for the actual parameter estimates that have already been obtained and whatever modifications to them an investigator may consider is worth examining. Such a procedure is unlikely to lead unequivocally to a clear choice of estimator or testing procedure in every case, but would allow an investigator to take cognizance of the apparent distributional characteristics of his statistical procedures in the precise setting of the model and the data set with which he is working. A pilot study inspired by this approach and based on the use of the Edgeworth approximation in the context of a simple dynamic simultaneous equations model is reported in Phillips [33]. Sections 5–7 of that paper detail the algebra involved in setting up the approximation, all of which can now be computerized including the complicated formulae for the cumulants of the sample moments of the data on which the statistics of interest depend. Moreover, these

formulae can be readily parameterized to allow for certain types of non-normality, possibly by the use of Gram Charlier type error distributions where the extent of the non-normality can be measured through the magnitude of certain critical parameters; and the effects of such non-normality on the distribution of interest will then show up in the tabulations or graphical plots that are output from the subroutine. Appendix E in Phillips [33] illustrates the use of this general procedure with a sequence of graphical plots showing the effect of a variety of parameter changes on the form of the finite sample distributions of two alternative estimators. More recent work on the computerized algebra of this approach in dynamic models has been done by Sargan and Tse [45].

If such work in small sample theory is to influence the way in which empirical work in econometrics is actually undertaken, particularly with respect to attempts to improve inferential accuracy, reduce bias, and advise on the merits of alternative estimators, then it is of paramount importance that the approximations that are being used in this way be already well tried in laboratory type conditions; their reliability in a variety of parameter environments and different regions of the distributions should be tested and recorded; their strengths and weaknesses should be known and then signposted when they are eventually incorporated into regression software as indicated in the last paragraph, perhaps in much the same way as users are, or should be, alerted to the possible pitfalls in the use of certain algorithms for optimization in non-linear estimation.

Work on the evaluation of small sample approximations is already well under way and some published studies do give valuable insight into such questions as the relative performance of different asymptotic

approximations and the reliability of the approximations in differing parameter environments. The approximations used in most of these studies have been based on three main methods: (i) the leading terms of Edgeworth-type asymptotic expansions of the distribution; (ii) normal approximations that use Nagar approximations to the first two moments; and (iii) the method of steepest descents leading to the saddlepoint approximation, whose use in statistics was first explored by Daniels [10, 11] and which has recently been the subject of renewed interest [12], [13], [14], [21], [31], [32].

Each of these methods of approximation is capable of representing the exact distribution of certain statistics to an acceptable degree of accuracy in certain parameter environments. This is confirmed by numerical evaluations in Anderson and Sawa [2, 3], Phillips [29, 31], and Holly and Phillips [21]. Moreover, the approximate distributions that have been obtained in the literature have already given valuable information concerning the small sample behavior of competing estimators and the adequacy of asymptotic theory in simple simultaneous equations and dynamic models. However, given the current state of our knowledge, the use of these methods in practical econometric work to advise on the choice of estimator and improve inferential accuracy is bound to encounter difficulties, some of them major.

First of all, there are certain parameter environments where the performance of the approximations is poor, sometimes a good deal worse than the asymptotic distribution (particularly in the case of the Edgeworth approximation). Unfortunately, the parameter environments for which this poor performance obtains are not at all unusual. As we might expect, given that the approximations are based on asymptotic series, this problem tends to become more widespread when sample sizes are small. Some

indication of the wayward nature of these approximations in certain parameter environments is already documented in [29] and [31]. Further examples are given later in this paper in the application of Section 8.

Secondly, although general formulae for the Edgeworth and Nagar approximations are now available ([29], [42]) and widely applicable, the saddlepoint technique is still only practicable in specialized cases where the characteristic function is available or simple integral formulae for the pdf can be used (such as in the case of ratios [11]) or where there exist a set of sufficient statistics for the parameters to be estimated [14]. No doubt progress will be made in tackling some of these latter difficulties but, in the meantime, they remain a barrier to the general use of the procedure in analyzing small sample distributions and influencing the way in which empirical work is actually undertaken.

Another difficulty that can arise in the use of the saddlepoint technique is that, for certain values of the argument of the pdf, singularities can occur within the strip of the imaginary axis containing the saddlepoint through which the path of integration is normally deformed. In such cases, this path of deformation is no longer permissible and special techniques must be used to smooth the approximation past the singularity; the resulting approximants are called uniform asymptotic expansions. Uniform approximants are typically much more complicated in form than the saddlepoint approximation (an example is given in [32]). They are not always easy to extract and further work is required to splice them with the saddlepoint approximation, where it does exist, to cover the whole of the distribution.

Finally, it seems difficult to embody additional information on the distribution in question into these approximations. To take a simple

example, in spite of the fact that the actual pdf is non-negative and the df monotonic it is sometimes awkward to modify the Edgeworth approximations so that they share these properties. To take a more complicated example, we often know or can find simply enough the leading term in the series representation of the exact pdf (in many cases, without knowing the full expression for the pdf). This leading term frequently has a simple algebraic form and is instrumental in determining the behavior of the exact distribution in certain domains, particularly the tails. Yet, even when this information is available, there seems to be no obvious or convenient way of building it into the Edgeworth, the Nagar or the saddlepoint approxitions. The resulting approximations, therefore, end up neglecting what is potentially very useful analytic information on the form of the distribution.

The above discussion may help to motivate what has been a major component of my own research strategy in small sample theory over the last five years and what now forms the major purpose of the present paper. This is to introduce a new technique of approximating distributions which is sufficiently general to be widely used and sufficiently accurate to be relied upon in empirical work. The technique is developed in very general terms and should be widely applicable in mathematical statistics and econometrics. It has the advantage, unlike the Edgeworth and saddlepoint approximations, of readily incorporating extraneous information on the distribution; even qualitative information or soft quantitative information such as that based on Monte Carlo experiments. Moreover, since the technique is not based on an asymptotic series expansion in terms of the sample size or concentration parameter, accurate approximations can be obtained even in very small samples. The technique should,

therefore, be most useful in cases where existing techniques based on asymptotic approximations run into difficulty.

The idea that underlies the new method is very simple. It is motivated by the observation that, in spite of the complex analytic forms of many of the exact pdf's presently known for econometric statistics, when we do turn around and obtain numerical tabulations or graphical plots of the densities we typically end up with well behaved, bounded, continuous functions that tend to zero at the limits of their domain of definition. The form of these pdf's strongly suggests that we should be able to get excellent approximations to them in the class of much simpler functions and certainly without the use of multiple infinite series. We need to deal with approximating functions (or approximants) that are capable of capturing the stylized form of a density: in particular, we will want the approximant to be able to go straight for long periods in a direction almost parallel to the horizontal axis and yet still be able to bend, quite sharply if necessary, to trace out the body of the distribution wherever it is located. One class of functions that seem particularly promising in this respect, as well as being simple in form, are rational functions. Even low degree rational functions can go straight for long periods and then bend quite sharply. In this, of course, they are very different from low degree polynomials whose graphs typically display a distinct roly poly character.

The present paper, which is envisaged as the first in a series on this topic, is concerned with the possibility of obtaining good global approximations to pdf's of a general class by means of rational functions and demonstrates a practical method of finding such an approximation in any given situation. In most cases, we have to accept that it will not

be possible to find the best approximant within a certain class (measuring best according to an already specified norm) without knowledge of the pdf itself. Obviously, such knowledge would partly defeat the purpose of finding an approximation; and, in what must be presently regarded as a very large number of important cases in econometrics, we are simply not in the position of possessing such knowledge. An important feature of the present article is the development of a technique for constructing good global approximants that makes use only of limited local information about the true pdf. The technique itself is based on the idea of working from local Taylor series approximations at certain points towards a global approximation which will perform well in the whole domain over which the distribution is defined, while retaining the good performance of the Taylor series approximations in the immediate locality of the points of expansion. This is, in part, achieved by the use of multiple-point Padé approximants which are rational functions constructed so as to preserve the local Taylor series behavior of the true pdf at certain points to as high an order as possible. The points selected for local expansion will often be simply the origin (in the central body of the distribution) and the tails. These local expansions can, in fact, be obtained from information about the characteristic function of the distribution so that direct knowledge even of the local behavior of the true pdf is not necessary for the application of the technique. The final step in the method is to modify the multiple point Padé approximant, to remove unwanted zeroes and poles which may occur in the bridging region between the points of local expansion.

The plan of the paper is as follows. The class of pdf's to be considered and the rational function approximants we will use are defined

in Section 2. A general theory of best uniform approximation in the context of density approximation by rational functions is then given in Section 3. This section gives existence, convergence, characterization and uniqueness theorems for rational approximants to density functions. But, since it is not critical to an understanding of many of the practical aspects of the approximation technique, this section may be omitted by those readers who are interested mainly in the nature, use and performance of the approximation. In a certain sense, Section 3 shows the best we can do with rational approximants without limiting the information that is available about the true pdf. The succeeding Sections 4, 5, 6 and 7 describe the method of constructing good global rational approximants to a given pdf by modifying multiple-point Padé approximants. These sections also give the general formulae needed in applications. In Section 8 the new method is applied to a simple simultaneous equations estimator, facilitating a comparison between the new and the existing techniques of approximation discussed earlier in this introduction. Some conclusions and generalizations of the study are presented in Section 9.

2. A GENERAL CLASS OF DENSITY FUNCTIONS AND RATIONAL APPROXIMANTS

To fix ideas, we write the estimator or test statistic in which we are interested as θ_T. In what follows, we treat θ_T as a scalar so that, when dealing with estimators, we are in effect concentrating on the marginal distribution of individual components of a complete vector of estimates. The characteristic function (cf) of θ_T is written as $cf(s) = E\left[e^{is\theta_T}\right]$ and is assumed to be absolutely integrable. This implies that θ_T has a bounded, continuous pdf given on inversion by

$$(1)\quad \mathrm{pdf}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-isx}\,\mathrm{cf}(s)\,ds\ .$$

Moreover, by the Riemann Lebesgue lemma, it follows from (1) that $\mathrm{pdf}(x) \to 0$ as $x \to \pm\infty$. Thus, the effect of the integrability requirement on $\mathrm{cf}(s)$ is to confine our attention to the class of densities covered by the following assumption:

ASSUMPTION 1. *θ_T has a continuous pdf which tends to zero at the limits of its domain of definition* $(\pm\infty)$.

Note that the boundedness of the pdf now follows from its continuity and behavior at $\pm\infty$. Assumption 1 covers a wide variety of densities arising in statistical and econometric work. It can, in fact, be extended to allow for certain types of discontinuity and singularity but this complicates the development of the approximants that follow. In this paper, we will keep to the class of densities defined by Assumption 1. This is sufficiently general to include all the usual simultaneous equations estimators and test statistics, as well as their extensions to models with lagged endogenous variables as regressors and autoregressive, moving average errors.

Having defined the class of density functions, the general problem of approximation takes the following form: for a particular density function $\mathrm{pdf}(x)$, find an approximating function which depends on a finite number of parameters whose values are selected in such a way that the approximating function is (in some sense) as close to the original density as possible over its entire domain of definition. Once stated in this way, it is clear that there are two major components to the problem. The first is the form the approximating function should take. The second is

the criterion of closeness of approximation to be used in selecting the best approximant. By a best approximant we mean the member (or members) of the given family of approximating functions whose closeness to the function pdf(x) cannot be improved by any other member of the same family. Thus, the second problem clearly raises the further question of whether or not there exists a best approximation to pdf(x) in the given family of approximants. This question of existence will be one of the issues addressed in the next section.

In selecting a suitable class of approximants, the requirements we need to take into account are largely dictated by the typical shape of a pdf and the interval over which the approximation is to be used. As discussed in the previous section, rational functions seem particularly promising in their capacity to capture this typical shape. They are also capable of providing good global approximations to an arbitrary continuous function over an infinite interval and, in this respect, are quite distinct from polynomials. In developing our theory it will be useful to make explicit the class of rational approximants we will be using as follows.

DEFINITION. *If* $s(x)$ *is a real continuous function satisfying* $s(x) > 0$ *and* $s(x) \to 0$ *as* $x \to \pm\infty$, *then we define the class of rational approximating functions by*

$$(2)\quad R_{m,n}(x;s,\gamma) = s(x)\frac{P_m(x)}{Q_n(x)}$$

$$= s(x)\frac{a_0 + a_1x + \ldots + a_mx^m}{b_0 + b_1x + \ldots + b_nx^n}, \quad -\infty < x < \infty, \quad m \leq n$$

where (i) the numerator and denominator are reduced to their lowest degree by the cancellation of identical factors;

(ii) m *and* n *are even integers with* $m \leq n$;

(iii) $\gamma' = (a_0, a_1, \ldots, a_n, b_0, b_1, \ldots, b_n) \in \Gamma$, *the parameter space, which is defined as the following subspace of* $n+m+2$ *dimensional Euclidean space*

$$\Gamma = \{\gamma: \sum_{i=0}^{n} b_i^2 = 1,\ Q_n(x) > 0 \ \textit{for all}\ x \in (-\infty,\infty)\} .$$

(iv) $R_{mn}(x;s,\gamma)/R_{mn}(-x;s,\gamma) \to 1$ *as* $x \to \infty$.

In what follows and where there is no risk of ambiguity we will simplify the representation of the rational fraction (2) by using the notation $R_{mn}(x) = R_{mn}(x;s,\gamma)$.

The coefficient function $s(x)$ in (2) is a vehicle by which additional information about the true density can be readily embodied in the approximant. This can be soft quantitative information, for example of the type that $pdf(x) > 0$ and $pdf(x) \to 0$ as $x \to \pm\infty$ (already explicit in $s(x)$); or hard quantitative information, for example of the type (i) that $pdf(x)$ has moments up to a certain order or (ii) that $pdf(x)$ takes an especially simple form in an important and relevant leading case. The role of $s(x)$ will be discussed in more detail in Section 6 and will be illustrated in the application of Section 8.

The condition $\sum_{i=0}^{n} b_i^2 = 1$ on the parameter space Γ is a normalization which eliminates the redundancy in the coefficients of the rational function (2). Other normalizations such as $b_0 = 1$ or $b_n = 1$ are possible and may be more useful in application. We will, in fact, later use the normalization $b_0 = 1$ in the application of Section 8 but the present definition of Γ is retained for the theoretical development.

The condition $Q_n(x) > 0$ ensures that the rational fractions (2)

have no poles on the real line and are compatible, therefore, with the class of density functions to be approximated. Since this is possible only when n is an even integer, we have introduced this requirement explicitly under (ii). On the other hand, if the density function we wish to approximate were non zero only on part rather than all of the real axis, it is clear that this requirement may be relaxed. Moreover, if a singularity in the density function is known to occur on the real axis, we could remove the condition $Q_n(x) > 0$. The position of the singularity may also be known, in which case this information can be incorporated directly into (2); otherwise it too must be approximated. In the following development under Assumption 1, we are working within the class of bounded continuous densities so that the problem of dealing with singularities does not occur. The Definition does not, however, exclude the possibility of approximants which possess zeroes and which become negative over part of the domain. This could, of course, be achieved directly by the further requirement $P_m(x) > 0$ parallel to the condition on $Q_n(x)$. But since $pdf(x) \to 0$ as $|x| \to \infty$ and in view of the error alternation property of the best approximant (to be discussed in the next section), such a requirement seems unnaturally restrictive, excluding as it does functions which become negative only on the extreme tails but which may nevertheless be excellent approximants to $pdf(x)$ over a wide domain.

The above Definition allows for rational fractions with numerator polynomials of any degree up to m and denominator polynomials of any (even) degree up to n. The normal approximant $R_{mn}(x;s,\gamma)$ in this class will involve polynomials $P_m(x)$ and $Q_n(x)$ of degrees m and n, respectively; as distinct from what is called an abnormal or degenerate

approximant in which the numerator and denominator polynomials are of lower than prescribed degree. In view of the known behavior of pdf(x) at infinity, we require in (iv) that, whatever the actual degrees of the numerator and denominator polynomials in $R_{mn}(x;s,\gamma)$, the approximant tends to the same limit at $\pm\infty$. The fact that m is an even integer at most equal to n (from (ii)) ensures that this is possible irrespective of the choice of the coefficient function s(x).

Later in this paper we will be concerned with the particular class of rational fractions (2) in which the numerator and denominator polynomials are of the same degree. This specialization of the theory is motivated by essentially practical considerations. First, the coefficient function s(x) will frequently be constructed so that it captures the behavior of the exact pdf(x) as x approaches the limits of its domain of definition. A rational fraction of equal degree is then immediately compatible with this behavior. Second, when the numerator and denominator are of equal degree, modifications to the coefficients that are designed to avoid unwanted zeroes and poles in the final approximant are easier to make. That this is of particular importance will be seen in Sections 6-7 where we develop a practical procedure for obtaining a good approximant of the type (2) when only limited information about pdf(x) is available. This procedure is based on the idea of modifying multiple-point Padé approximants, which in crude form will frequently possess zeroes and poles that need to be removed in order to improve the approximation over the whole real line. Finally, numerical experience with rational function approximations in applied mathematics (see, for example, [19]) suggests that rational fractions with numerator and denominator of equal or near equal degree tend, on the whole, to give better approximations than

those for which the degrees differ markedly. Taking an extreme case of comparison, polynomial approximations usually become unsatisfactory when it is necessary to approximate a function over a wide interval. In particular, the Weierstrass Theorem is invalid on an infinite interval. Moreover, polynomials lack the capacity to turn corners sharply and then go straight for long periods, particularly in a direction almost parallel to the horizontal axis. These properties are useful ones for a density function approximant to be capable of capturing as we discussed in the Introduction. An important feature of rational fraction approximations is that even low degree fractions of the type (2) are flexible enough to assume this behavior. This is endorsed by the large number of numerical results with rational approximants reported by Hastings [20] and Hart [19]. It will also be confirmed in our own application of the technique reported in Section 8.

In order to develop a theory for the goodness of approximation based on members of the class (2) we introduce a norm to measure the error in the approximation. We will use the uniform norm (also known as the Tchebycheff or L_∞ norm) defined as

$$(3) \quad \|f(x)\| = \sup_{x\in(-\infty,\infty)} |f(x)| \ .$$

If we now let $f(x) = pdf(x) - R_{m,n}(x;s,\gamma)$ denote the approximation error, our problem is, for given value of m, n and a given function $s(x)$, to find a value of γ which minimizes the maximum error. At this value of γ, $R_{m,n}(x;s,\gamma)$ is then called a best uniform (or Tchebycheff) approximation to $pdf(x)$.

Other choices of norm are certainly possible and will generally lead to different best approximations, where they exist. However, for

accurately approximating pdf(x) over a wide interval the choice of the uniform norm seems most appropriate. Other norms are clearly more appropriate for best approximation of the df or the moments of the distribution, at least if our point of departure is the density function itself. An alternative approach which seems preferable is to work with these functions directly and retain the uniform norm to evaluate their rational approximants. We might also consider a measure based on the relative error of the approximant, which would lead us to use the error function $[pdf(x)^{-1}\{pdf(x) - R_{m,n}(x;s\gamma)\}]$ in evaluating the approximant. These are issues that will be explored in later work.

3. BEST UNIFORM APPROXIMATION BY RATIONAL FUNCTIONS

The theory of best uniform approximation of real continuous functions by rational fractions has a long history. One of the earliest investigations was undertaken by Tchebycheff [47]. Frobenius [18] and Padé [28] both systematically explored the properties of a specialized class of rational approximants now known as Padé approximants, which we will consider in Sections 5 and 6 below. In the complex domain, Runge [40] (see Rudin [39], chapter 13) established the possibility of uniform approximation of analytic functions by rational fractions with preassigned poles. A general theory of approximation in the complex domain by rational functions was developed in the treatise by Walsh [50]. Extensive modern treatments of the subject covering all the classical results on the approximation of real valued functions are given in the volumes by Rice [37] and Meinardus [25].

The theory in this literature has been primarily, but not exclusively, concerned with the approximation of functions defined over compact sets.

This theory does not always admit a trivial extension to unbounded regions. Some results, notably the Weierstrass Theorem on polynomial approximation, are false unless the domain of approximation is finite. Other results, such as the Tchebycheff equioscillation theorem for the error on a rational approximant (Theorem 3 below) remain only partially valid and their proofs undergo essential alterations to accommodate the fact that the domain of approximation is an infinite interval.

In its general form, the problem of approximating a pdf by a rational fraction in the class (2) is defined over the infinite interval $(-\infty,\infty)$. Under Assumption 1, $pdf(x) \to 0$ as $|x| \to \infty$, so that if we add the point at infinity ∞ to the real axis $pdf(x)$ is continuous on the extended real axis $[-\infty,\infty]$. In this form the problem comes within the framework of rational approximation in $C[-\infty,\infty]$, the space of all continuous functions $f(x)$ over $(-\infty,\infty)$ for which $\lim_{x\to\infty} f(x) = \lim_{x\to-\infty} f(x)$ and the limit is finite. Some aspects of rational approximation in $C[-\infty,\infty]$ have received attention in the work of Walsh [49], [50], Achieser [1] and Timan [48]. While working primarily with analytic and meromorphic functions of a complex variable, Walsh in both [49] and [50] proved an existence theorem for best uniform approximation by rational functions with numerator and denominator polynomials of equal degree. This theorem holds for an arbitrary set in the extended complex plane under the condition that it be dense in itself. As such, it certainly applies to real valued functions in $C[-\infty,\infty]$, but it does not apply directly to the class (2) involving a general coefficient function $s(x)$. Walsh does not give characterization and uniqueness theorems for the best approximant. In fact, he shows the converse of the latter, that the best approximant may fail to be unique for certain functions and domains in the complex plane.

The last two references by Achieser [1] and Timan [48] deal explicitly

with real valued functions. Achieser [1], in particular, develops fully a theory of uniform approximation applicable to a restricted version of our class (2). Achieser's restriction is on the class of admissable coefficient functions $s(x)$. He requires that

$$(4) \quad \lim_{x\to\infty} s(x)x^k = \lim_{x\to-\infty} s(x)x^k \neq 0$$

where $k = m-n$ (in our notation). In our context, (4) is too restrictive. We may, for example, wish to impose rather different behavior on $s(x)$ at infinity. Thus, if we know that $pdf(x) = O(|x|^{-\mu})$ $\mu > 0$, or $O(e^{-\alpha|x|^{\beta}})$ $\alpha,\beta > 0$, as $|x| \to \infty$, it makes sense to incorporate similar behavior in the coefficient function $s(x)$ and then set $m = n$ in the rational fraction (2). These possibilities are excluded by (4). Moreover, (4) provides an essential simplification of the theory in that ∞ becomes an allowable extreme point in the approximation error. Under this simplification the theory of best uniform approximation in $C[-\infty,\infty]$ is essentially the same as it is for $C[a,b]$ over a finite interval $[a,b]$. In particular, all the major results apply as for a finite interval. When we relax (4) and require only that $s(x) > 0$ and $s(x) \to 0$ for $|x| \to \infty$ as in our Definition of (2), this is no longer the case. While the existence, uniqueness and convergence results apply as for finite intervals, the usual characterization theorem (Achieser [1], p. 55, Meinardus [25], pp. 161-162, Rice [37], p. 80) fails. This theorem tells us that a necessary and sufficient condition for the best uniform approximant is that its error curve oscillate a specific number of times, which depends on the degree of the approximant. Theorem 3 below shows that for rational approximation of $pdf(x)$ over $[-\infty,\infty]$ within the class (2),

this condition is sufficient but not necessary.

Thus, while the work of Walsh and Achieser, that we have just discussed is very relevant to our problem of best approximation to pdf(x) within the class of approximants (2), we cannot completely rely on this work in the development of our own theory. Moreover, with regard to the important problem of characterizing the best approximant, our theory involves an important departure from existing results. The following theorems form the basis of this theory. Their proofs, together with some subsidiary technical material, are given in the Appendix in order to facilitate the reading of the paper by those interested in its main ideas.

THEOREM 1 (Existence). *If* pdf(x) *satisfies Assumption 1 on* $(-\infty,\infty)$, *then there exists a best uniform approximant to* pdf(x) *in the class of rational functions defined by* (2).

This theorem ensures that, given pdf(x), there exists a set of parameters γ^* and a corresponding rational function $R'_{nn}(x;s,\gamma^*)$ in the class (2) for which

$$(5) \quad \left\| \text{pdf}(x) - R'_{mn}(x;s,\gamma^*) \right\| = \inf_{\gamma\in\Gamma} \left\| \text{pdf}(x) - R_{mn}(x;s,\gamma) \right\| .$$

The prime in $R'_{mn}(x)$ is used to distinguish this rational fraction from a limiting function $R_{mn}(x;s,\gamma^*)$ which does not necessarily belong to the class (2) (it may not even be continuous) because the set Γ is not closed. This problem is discussed together with an example in the Appendix. It does not affect the conclusion of Theorem 1 that there does indeed exist a rational function belonging to the class (2) which is the closest possible approximant to pdf(x) in the sense of the uniform norm (3).

THEOREM 2 (Convergence). *If* pdf(x) *satisfies Assumption 1 on* $(-\infty,\infty)$, *if* pdf(x)/s(x) *has a finite limit or tends to infinity as* $x \to \pm\infty$, *and if*

$$(6) \quad E(n;s) = \| \text{pdf}(x) - R'_{nn}(x;s,\gamma^*) \|$$

where $R'_{nn}(x;s,\gamma^*)$ *is the best uniform approximant to* pdf(x) *in the class of rational fractions defined by* (2) *with* $m = n$, *then*

$$(7) \quad \lim_{n\to\infty} E(n;s) = 0 .$$

As the degree of the best approximating rational function increases the error $\| \text{pdf}(x) - R'_{mn}(x;s,\gamma^*) \|$ must be at least as small. Theorem 2 shows that the best approximant $R'_{nn}(x;s,\gamma^*)$ converges to pdf(x) uniformly as $n \to \infty$. It follows that, for any choice of density function satisfying Assumption 1, there is an arbitrarily close rational approximant. This theorem does not, however, give an order of magnitude on the error of approximation as $n \to \infty$. Some results on error magnitudes and rates of convergence are available in the approximation theory literature but these apply in the main to special functions. To take two examples: (i) it is known (Newman [26]) that whereas the error on an n^{th} degree polynomial approximation to $|x|$ over $[-1,1]$ has order n^{-1}, the error on a rational approximant of type $R_{nn}(x;\ s(x) = 1)$ has order at most $\exp(-n^{1/2})$; (ii) it is also known [17] that the error on a rational approximant of type $R_{nn}(x;\ s(x) = 1)$ to the function $\exp(-|x|)$ over $(-\infty,\infty)$ has order at most $\exp(-\alpha n^{1/2})$ for some constant $\alpha > 0$. Both of these rather specialized results on rational approximants suggest that, at least for certain classes of functions, the rate of convergence of polynomial approximants can be dramatically improved by the use of rational

functions with numerator and denominator of equal degree. These results corroborate the numerical experience described in Section 2 which suggests that rational fractions of this type tend on the whole to provide better approximations than those for which the numerator and denominator polynomials differ markedly in degree. In this paper, we will concentrate on rational fractions of this type $R_{nn}(x;s,\gamma)$ in the development of our practical procedure based on modified Padé approximants in Sections 5-8.

THEOREM 3 (Error Bound). *Let the rational function of the class* (2)

$$(8)\quad R_{mn}(x;s,\gamma) = s(x)\frac{P_m(x)}{Q_n(x)} = s(x)\frac{a_0 + a_1x + \ldots + a_{m-\mu}x^{m-\mu}}{b_0 + b_1x + \ldots + b_{n-\nu}x^{n-\nu}}$$

be in its lowest terms with no common factors in $P_m(x)$ *and* $Q_n(x)$, *with* $a_{m-\mu} \neq 0$, $b_{n-\nu} \neq 0$ *and* $0 \leq \mu \leq m$, $0 \leq \nu \leq n$. *We set* $d = \min(\mu,\nu)$ *and* $N = m+n-d+2$. *If at the consecutive points* $x_1 < x_2 < \ldots < x_N$ *in the interval* $[-\infty,\infty]$ *we have*

$$(9)\quad \mathrm{pdf}(x_i) - R_{mn}(x_i) = (-1)^i\lambda_i$$

where all λ_i *have the same sign and are all different from zero, then*

$$(10)\quad \inf_{\gamma\in\Gamma} \|\mathrm{pdf}(x) - R_{mn}(x;s,\gamma)\| \geq \min\{|\lambda_1|, |\lambda_2|, \ldots, |\lambda_N|\}\ .$$

The same result is true if $R_{mn}(x;s,\gamma) \equiv 0$ *in* (8), *in which case we take* $N = m+2$.

This theorem can be used to find a lower bound for the deviation of the best approximation. It suggests that when the error curve oscillates a sufficient number of times (which will usually be $N = n+m+2$,

with the "defect" in the rational fraction $d = 0$) and the extreme values of the oscillation are close in absolute value, the rational fraction is also close to the best approximant in that class. When the error curve oscillations differ markedly in magnitude, there will be considerable scope for improvement in the approximant.

THEOREM 4 (Tchebycheff's Equioscillation Theorem). *Suppose* $R_{mn}(x;s,\gamma)$ *is a rational fraction of the class* (2), *having the same form as* (8) *above. Let* $d = \min(\mu,\nu)$ *as in Theorem 3 and let*

$$(11) \quad h = \|\mathrm{pdf}(x) - R_{mn}(x;s,\gamma)\| > 0 \ .$$

Part A: Necessary Conditions

If $R_{mn}(x;s,\gamma)$ *is the best uniform approximant to* $\mathrm{pdf}(x)$ *then either*

(i) $d = \nu$ *(numerator most degenerate) and the number of alternations in the error*

$$(12) \quad e(x) = \mathrm{pdf}(x) - R_{mn}(x;s,\gamma)$$

(i.e., the number of consecutive points of the interval $[-\infty,\infty]$ *at which the error* $e(x)$ *takes on its maximum value* h *with alternate changes of sign) is at least* $N = n + m - d + 2$;

or (ii) $d = \mu$ *(denominator most degenerate) and the number of alternations of* $e(x)$ *is at least* $N = n+m-d$;

or (iii) *if* $R_{mn}(x) \equiv 0$ *then the number of alternations of* $e(x)$ *is at least* $N = m+2$.

Part B: Sufficient Conditions

If $R_{mn}(x;s,\gamma)$ *has the form* (8) *and the number of alternations of the error* $e(x)$ *is at least* $N = n+m-d+2$ *then* $R_{mn}(x;s,\gamma)$ *is the best uniform approximant to* $pdf(x)$ *in the class* (2).

This result characterizes the best approximant and formalizes the discussion following Theorem 3. Specifically, Part B tells us that if the error curve takes on its maximum value with alternate changes of sign at least $N = n+m-d+2$ times then this is sufficient to ensure that the rational function (8) is a best approximant. In the classical characterization theorem (Achieser [1], pp. 55-57, Meinardus [25], pp. 161-162), this condition is also necessary. It is an interesting feature of the present problem that this condition is no longer always necessary. This difference with the classical theory only occurs when the best approximant has a degenerate denominator as in Part A(ii) (and even then not always). In the usual non-degenerate situation where the best approximant in the class (2) has numerator and denominator polynomials of full degree (with $\mu = \nu = d = 0$) the standard alternation theorem applies with the number of alternations being at least $N = n+m+2$.

To show that the difference between Theorem 4 and the classical theory is not trivial, we construct the following example. Let

$$(13)\quad pdf(x) = \frac{1}{\pi}\left(\frac{1}{1+x^2}\right) + t(x)\ , \quad -\infty < x < \infty$$

where $t(x)$ is zero except over three intervals $[-1-\eta, -1+\eta]$, $[-\varepsilon,\varepsilon]$ $[1-\eta, 1+\eta]$. We define $t(x)$ to be piecewise linear and continuous over $-\infty < x < \infty$, to take on the value zero at the ends of the three intervals just defined and, to provide spikes in the function in each of these

intervals, we set $t(-1) = +h$, $t(0) = -h$, $t(1) = +h$ where h is positive (and less than $1/\pi$). We set $\eta = \varepsilon/2$ to normalize the area under (13). The graph of (13) is displayed in Figure 1. Consider the approximant to (13) within the class (2) with $m = 0$, $n = 2$, $s(x) = [\pi(1+x^2)]^{-1}$ and of the form

$$(14) \quad R_{02}(x;s,\gamma) = \left[\frac{1}{\pi(1+x^2)}\right]\frac{a_0}{1+b_2x^2} , \quad -\infty < x < \infty$$

The best approximant to (13) of the form (14) is obtained with $a_0 = 1$, $b_2 = 0$ and we have $R'_{02}(x;s,\gamma^*) = s(x) = [\pi(1+x^2)]^{-1} = R'_{00}(x;s,\gamma^*)$, the same fraction as for the class in which $m = n = 0$. The error on the approximant is $e(x) = t(x)$ and an alternant occurs at the points -1, 0, $+1$ with $e(-1) = +h$, $e(0) = -h$, $e(1) = +h$. Note that the number of alternations here is 3 which is less than the number $n+m+2-d = 4$ which would have been required if the classical theory applied.

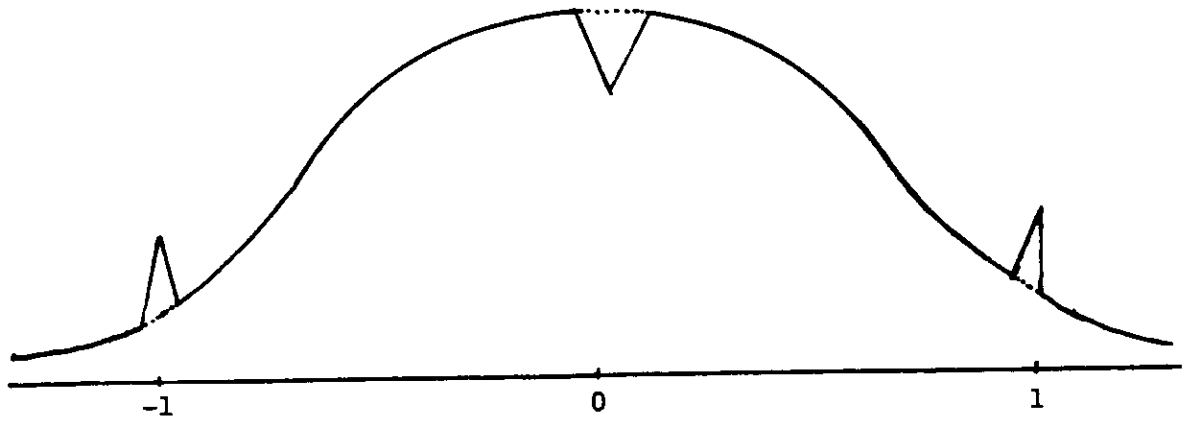

FIGURE 1

pdf(x) (——) and best approximant (·····)

To see that $R'_{00}(x) = [\pi(1+x^2)]^{-1}$ is indeed the best approximant of the form (14) we note that the error can be reduced at $x = 0$ only by setting $a_0 < 1$. Now since (14) must approximate pdf(x) over the whole real axis we require $b_2 \geq 0$ so that reducing the error at $x = 0$ must always increase the error at $x = \pm 1$.

The same result holds for the approximant

$$(15) \quad R_{02}(x;s,\gamma) = \left[\frac{1}{\pi(1+x^2)}\right] \frac{a_0}{1 + b_1 x + b_2 x^2}, \quad -\infty < x < \infty$$

In order to reduce the error on the approximant $R'_{00}(x) = s(x) = [\pi(1+x^2)]^{-1}$ at $x = 0$ we need to set $a_0 < 1$. As before b_2 must be non negative and this means that the error on (15) must be greater than h at either -1 or $+1$ depending on whether b_1 is negative or positive.

This example proves that the classical theory no longer applies in our problem. Note that the degeneracy in the approximant in this case occurs in the denominator. There is no degeneracy in the numerator and we have $m = 0$, $n = 2$, $d = \min(\mu,\nu) = 0$ so that the necessary number of alternations in the classical theory is $N = n+m+2-d = 4$, while the actual number of alternations in the best approximant is 3.

Finally, we have the following result which establishes the uniqueness of the best approximant in the rational family (2).

THEOREM 5 (Uniqueness). *The best uniform approximant to* pdf(x) *in the class of rational fractions defined by* (2) *is unique when reduced to its lowest terms.*

4. LOCAL EXPANSIONS FOR DENSITIES AND PADÉ APPROXIMANTS

The theory of the last section shows that for a given pdf in the class defined by Assumption 1 there exists a best rational fraction approximant of the type (2) and that, as we increase the degree of the approximant, this converges to pdf(x) over the entire real axis. In any practical situation, of course, we will need to prescribe the degree of the approximant to be used and attempt to find the best approximant in the given class. This normally requires numerical methods and the algorithms discussed in the literature[1] rely on knowledge of the true function values at a grid of points as well as, in certain cases, the function derivatives. This seems too much to expect in an econometric context where, even in those cases where the exact density function is known in analytic form, numerical computations are often impossible because of convergence problems with the multiple series representation of the density or the inadequate tabulations of the special polynomials that appear in the analytic expressions.

We are, therefore, left with the problem of how, in a given situation, to get close to the best approximant in the class (2) without having to rely on arbitrary evaluations of the exact distribution. The solution we present to this problem in the present and succeeding Sections of the paper is based on the idea of using the local behavior of the true density in the body of the distribution and in the tails to construct a global approximation of the form (2). In principle, the procedure we develop for moving from local to global density approximations can be based on

---

[1] See for example [25], pp. 170-171 and [37], Ch. 6.

the knowledge of local behavior at an arbitrary set of points. But, in practice, it will often be sufficient to use information concerning the local behavior of the density in the tails and around the center of the distribution. The application we consider in Section 8 will show that this information is sufficient to secure excellent global approximations to rather complicated density functions even with rational fractions of low degree.

Local behavior of density functions can take the form of expansions about the value of the function at a certain point or perhaps estimates of the function values obtained from Monte Carlo simulations. We will deal with the case where some analytic information from local expansions is available while, at the same time, it should be clear how the procedure we develop can also be used to accommodate Monte Carlo evidence. Thus, the procedure has the advantage that it can utilize soft as well as hard quantitative information that may be available about the true density.

Our present analytic knowledge of the exact distribution of a variety of econometric estimators and test statistics shows that there exists an asymptotic expansion of the density function in ascending powers of x^{-1} as the argument x approaches the limits of its domain of definition $(\pm\infty)$. In general, we can write the expansion about infinity in the form

$$(16) \quad \text{pdf}(x) \sim t(|x|)\{\alpha_0 + \alpha_1/x + \alpha_2/x^2 + \alpha_3/x^3 + \alpha_4/x^4 \ \ldots\ \}$$

as $x \to \pm\infty$. The coefficient function $t(|x|) \to 0$ as $|x| \to \infty$ and, in the case of most of the common simultaneous equations estimators, is of the form $t(|x|) = |x|^{-k}$ where $k \geq 2$. Thus, in the case of the two stage least squares estimator $k = \ell+2$ where ℓ is the degree of over-

identification in the equation being estimated. An expansion of the type (16) was developed by Sargan and Mikhail [44] for the instrumental variable estimator and was used by Sargan [43] in the analysis of Monte Carlo estimates of moments that do not exist.

At points $\{d_i : i = 1, \ldots, I\}$ where $pdf(x)$ is continuously differentiable to an appropriate order we have the Taylor expansions

$$(17) \quad pdf(x) = \beta_{i0} + \beta_{i1}(x-d_i) + \beta_{i2}(x-d_i)^2 + \beta_{i3}(x-d_i)^3 + \beta_{i4}(x-d_i)^4 + \ldots \qquad (i = 1, \ldots, I)$$

In a number of cases, we also have the analytic form of the leading term in the series representation of the density. If we denote this leading term by $w(x)$, then it will be useful to consider extensions of the expansion (17) which take the form

$$(18) \quad pdf(x) = w(x)\{\beta_{i0} + \beta_{i1}(x-d_i) + \beta_{i2}(x-d_i)^2 + \beta_{i3}(x-d_i)^3 + \beta_{i4}(x-d_i)^4 + \ldots\}$$

Two obvious choices of the points d_i are: (i) the origin, particularly for certain test statistics like the "t" ratio; and (ii) the true value of the relevant parameter, when $pdf(x)$ refers to the marginal distribution of a certain estimator; or perhaps its limit in probability when dealing with an inconsistent estimator.

Although expansions such as (16) and (17) usually produce good approximations only in the immediate neighborhood of the point of expansion, they can be used to construct approximations which perform well outside the immediate locality of the approximation, while retaining the good behavior of the original expansions with the locality. With reference to (17), the fourth degree polynomial in x may yield a good approximation

to pdf(x) in a neighborhood of the point d_i ; but, in most cases, its performance will rapidly deteriorate outside of this neighborhood and it will be quite inadequate as an approximation on the tails. On the other hand, the coefficients β_{ij} in the expansion (17) usually contain information which can produce greatly improved approximations outside the range in which the expansion (17) itself is immediately useful. That this is so is demonstrated by the extensive practical experience with Padé approximants in the applied mathematics literature. These approximants are rational fractions for which the corresponding Taylor series matches the Taylor series expansion of a given function to as many powers as is possible. The following example was used by Baker [5]:

$$(19)\quad f(x) = \left(\frac{1+2x}{1+x}\right)^{1/2} = 1 + \frac{1}{2}x - \frac{5}{8}x^2 + \frac{13}{16}x^3 - \frac{141}{128}x^4 + \ldots$$

The Taylor series for f(x) in (19) has radius of convergence equal to 1/2. Yet as x becomes large f(x) is a well behaved function which tends to $\sqrt{2}$ as $x \to \infty$. Using only the first three coefficients 1, $\frac{1}{2}$, $-\frac{5}{8}$ in (18), we construct the Pade approximant

$$(20)\quad \frac{1+(7/4)x}{1+(5/4)x} = 1 + \frac{1}{2}x - \frac{5}{8}x^2 + \frac{25}{32}x^3 - \ldots$$

This has the same Taylor series expansion about the origin as f(x) to $O(x^2)$ and it tends to 7/5 = 1.4 as $x \to \infty$. Thus, using only three coefficients in a local expansion about the origin, the Padé approximation (19) provides an approximation at infinity to f(x) which differs at the second decimal place. Even within the radius of convergence of the Taylor expansion (20) outperforms the Taylor expansion: for instance, at $x = 1/4$, $f(x) = 1.0954451$, the first three terms of the Taylor

expansion give 1.0859375 while (20) equals 1.0952381 providing at least another decimal place of accuracy.

This example suggests that Padé approximants can have the useful property of accelerating the convergence of a given power series within its circle of convergence, while at the same time considerably extending the domain over which truncated series expansions can give useful results. These features make Padé approximants attractive for constructing first step rational fraction approximations from the information embodied in purely local density expansions such as (16), (17), or (18). Section 6 will be devoted to the algebraic details of this construction and will give the appropriate formulae.

5. USE OF THE CHARACTERISTIC FUNCTION IN DETERMINING LOCAL DENSITY EXPANSIONS

In order to construct rational approximants from local density expansions, the coefficients in these expansions need to be available up to a certain order. When the true density is known and computable this will not, of course, present any difficulty. But, in the most usual situation, the true density is not known in analytic form. In many of these cases we may still be able to find analytically the leading term in the (series expression for the) density, under an appropriate null hypothesis; and tail expansions up to a limited extent may also be available. In these latter cases the procedure developed in Sections 6 and 7 can be applied directly. In the remaining cases, an alternative approach is required. This section of the paper shows that local density expansions such as (16) and (17) can be extracted from knowledge of the characteristic function and not the density. Theorem 8 shows in particular,

that local information about the behavior of the characteristic function in the neighborhood of the origin is by itself sufficient to develop the tail expansion (16) for the density pdf(x) .

We start with the following two basic results which relate the tail behavior of density functions to the regularity properties of the characteristic function. They follow without difficulty from the standard discussions on this subject in the literature;[1] but they also demonstrate that we need to go somewhat further to extract a tail expansion of the form (16).

THEOREM 6. *If the distribution with density* pdf(x) *and characteristic function* cf(s) *has finite* $(M-1)^{th}$ *absolute moment, then* cf(s) *is* M-1 *times continuously differentiable and the derivatives* $cf^{(n)}(s) \to 0$ *as* $s \to \pm\infty$ *for each* $n = 0, 1, \ldots, M-1$.

THEOREM 7. *If* cf(s) *is* M-1 *times continuously differentiable, if* $cf^{(n)}(s) \to 0$ *as* $s \to \pm\infty$ *and if* $cf^{(n)}(s)$ *is absolutely integrable for each* $n = 0, 1, \ldots, M-1$, *then*

$$pdf(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-isx} cf(s)ds = \frac{(ix)^{-M+1}}{2\pi}\int_{-\infty}^{\infty} e^{-isx} cf^{(M-1)}(s)ds = O(x^{-M+1})$$

as $x \to \infty$.

Theorem 7 shows that $pdf(x) \to 0$ as $x \to \pm\infty$ at least as fast as $|x|^{-M+1}$; but this is, in general, not a very sharp result. For, if the $(M-1)^{th}$ absolute moment of the distribution exists and pdf(x) satisfies Assumption 1 then we would expect that $pdf(x) = O(x^{-M-\delta})$ for some $\delta > 0$. For example, in the case of the Cauchy distribution

[1] See, for example, Feller [16] or Lukacs [24].

$cf(s) = e^{-|s|}$ and Theorem 7 demonstrates that $pdf(x) = O(1)$ whereas, in fact, $pdf(x) = O(x^{-2})$. Thus, Theorems 6 and 7 are not very helpful in providing local expansions about infinity of the form (16).

However, a sharper result which does lead directly to the asymptotic expansion (16) can be obtained from a more explicit representation of the behavior of the characteristic function as the following Theorem shows.

THEOREM 8. *Let the characteristic function* $cf(s)$ *be absolutely integrable and its behavior as* $s \to 0$ *be given by the asymptotic series*

$$(21)\quad cf(s) \sim e^{i\eta s}\left\{\sum_{m=0}^{M-1} p_m (is)^m + |s|^{\mu} \sum_{j=0}^{\infty} \sum_{k=0}^{K(j)} \sum_{\ell=0}^{L(j)} q_{jk\ell} |s|^{\nu j} (i\,\mathrm{sgn}(s))^k (\ell n|s|)^{\ell}\right\}$$

where η, μ, ν, p_m, $q_{jk\ell}$ *are real constants and* $\mathrm{sgn}(s) = 1, 0, -1$ *for* $s > 0, = 0, < 0$. *Then the corresponding probability density function* $pdf(x)$ *has the following asymptotic expansion as* $|x| \to \infty$

$$(22)\quad pdf(x) \sim \frac{1}{\pi |x-\eta|^{\mu+1}} \sum_{j=0}^{\infty} \sum_{k=0}^{K(j)} \left[\sum_{\ell=0}^{L(j)} (q_{jk\ell}\, \partial^{\ell}/\partial z^{\ell}) \Gamma(z+\mu+1) |y|^{-z} \cdot \frac{1}{2}\left\{ i^k e^{-\frac{1}{2} i\pi\, \mathrm{sgn}(y)(z+\mu+1)} + (-i)^k e^{\frac{1}{2}\pi\, \mathrm{sgn}(y)(z+\mu+1)} \right\}\right]_{\substack{z=j\nu \\ y=x-\eta}}$$

The representation (21) is sufficiently general to include a very wide class of distributions and should cover most distributions of practical interest in mathematical statistics. In general, we will find in most applications of this Theorem that $\mu \geq M$, $\nu > 0$, $K(j) = 0$ and $L(j) = 0$ or 1 for all j. The first component in braces on the right side of (21) is analytic and ensures, when $\mu \geq M$, that integral moments of the

distribution will exist to order $M-1$ if this is an even integer and to order $M-2$ if $M-1$ is odd [24]. In cases where M is finite and the distribution does not possess all its moments, the second component of (21) is important in the local behavior of $cf(s)$ in the locality of the origin and is instrumental in determining the form of the tails of the distribution as the result (22) shows.

To illustrate the use of Theorem 8, we go back to the example of the Cauchy distribution with $cf(s) = e^{-|s|}$. In this case

$$cf(s) \sim \sum_{j=0}^{\infty} \frac{(-1)^j |s|^j}{j!}$$

and we deduce from (22) by setting $\mu = 1$, $\eta = 0$, $\nu = 1$, $K(j) = 0$ and $L(j) = 0$ for all j that

$$pdf(x) \sim \frac{1}{\pi x^2} \sum_{j=0}^{\infty} \frac{(-1)^{j+1}\Gamma(j+2)\cos\left\{\frac{1}{2}\pi(j+2)\right\}}{(j+1)!} |x|^{-j}$$

$$= \frac{1}{\pi x^2} \sum_{n=0}^{\infty} (-1)^n (x^2)^{-n}$$

where $2n = j$. This expansion can be verified directly from the probability density $pdf(x) = [\pi(1+x^2)]^{-1}$ itself.

THEOREM 9. *If* $s^N cf(s)$ *is absolutely integrable, then the local expansion of* $pdf(x)$ *about the point* $x = d_i$ *is given by*

$$(23) \quad pdf(x) = \sum_{j=0}^{N-1} \beta_{ij} (x - d_i)^j + O(x - d_i)^N$$

where

$$\beta_{ij} = \frac{1}{2\pi j!} \int_{-\infty}^{\infty} (-is)^j e^{-isd_i} cf(s) ds , \quad j = 0, 1, \ldots, N-1 .$$

Local expansions of the type discussed in this section for the tails and the body of the distribution can also be extracted under similar conditions for the distribution function. These expansions will be useful in the development of a corresponding theory of global approximation for the distribution function rather than the density and will be discussed in a later paper. Expansions of this type are already given for many of the common distributions in the statistical literature ([51]).

6. MULTIPLE-POINT PADÉ APPROXIMANTS

As discussed in Section 4, Padé approximants can be used to improve the convergence properties of local Taylor expansions and have the useful additional property that they frequently extend the domain over which these local expansions provide good approximations. This section will show how Padé approximants can be derived from local density expansions such, as (16) and (17). These approximants will provide a preliminary set of rational fractions. They can then be used directly as approximations to $pdf(x)$ or modified so that they belong to the class of rational fractions (2) and have satisfactory global behavior. The question of modifying the preliminary rational fractions will be taken up in Section 7.

We start by writing the density function in the form

$$(24) \quad pdf(x) = s(x)pdf_s(x) , \quad -\infty < x < \infty$$

where $s(x)$ is a real continuous function satisfying $s(x) > 0$ over the entire real axis and $s(x) \to 0$ as $x \to \pm\infty$. This representation of $pdf(x)$ reconciles with the class of rational fractions defined in (2) and allows us to accommodate information about the coefficient functions $t(|x|)$ and $w(x)$ which appear in the local density expansions (16) and

(18). In many cases, $s(x)$ will represent the leading term in the multiple series representation of the density $pdf(x)$ and, in such cases, $s(x)$ will usually be identical to $t(|x|)$ and $w(x)$. When this leading term in the density is unknown, a suitable alternative will be to set $s(x) = t(|x|)$ directly or some modified form of $t(|x|)$ which has the same asymptotic behavior but which is well behaved elsewhere on the real axis. If necessary, the expansion (18) can then be adjusted to take account of this modification so that (17) is correct to the same number of terms.

If we write the local expansions of $pdf_s(x)$ in the form

$$(25)\quad pdf_s(x) \sim \alpha_0 + \alpha_1 x^{-1} + \alpha_2 x^{-2} + \alpha_3 x^{-3} + \alpha_4 x^{-4} + \dots , \quad x \to \pm\infty$$

$$(26)\quad pdf_s(x) \sim \beta_{i0} + \beta_{i1}(x-d_i) + \beta_{i2}(x-d_i)^2 + \beta_{i3}(x-d_i)^3 + \beta_{i4}(x-d_i)^4 + \dots ,$$
$$(i = 1, \dots, I) , \quad x \to d_i$$

our problem is to construct a rational fraction of the form

$$(27)\quad [n/n] = \frac{P_n(x)}{Q_n(x)} = \frac{a_0 + a_1 x + \dots + a_n x^n}{b_0 + b_1 x + \dots + b_n x^n} , \quad n = \text{an even integer}$$

which has the same local behavior as (25) and (26) and to as high an order as possible. Such a rational fraction is called a multiple-point Padé approximant and is discussed by Baker [5, Ch. 8].[1] Our choice of the Padé approximant $[n/n]$ in which both numerator and denominator polynomials have the same degree is motivated by the considerations that were

[1] This work together with [6] and the collection of articles in [41] provide a systematic coverage of the extensive literature on the theory of Padé approximants and their applications, particularly in mathematical physics.

discussed in detail in Section 2 and in the remarks following Theorem 2 in Section 3.

The equations which define (27) can best be introduced by considering the approximant based on the Taylor series about a single point. We take the case of (26) with $d_i = 0$ and normalize (27) by setting $b_0 = 1$. This normalization ensures that $Q_n(0) = 1 > 0$ so that the [n/n] approximant will not have a pole at the origin (this now being the point of expansion of the Taylor series (26). The coefficients of [n/n] are now determined by the equation

$$(28) \quad pdf_s(x)Q_n(x) - P_n(x) = O(x^{2n+1})$$

Explicitly, we have the relations

$$
\begin{aligned}
&\beta_{i0} && && && && && && = a_0 \\
&\beta_{i1} && + \beta_{i0}b_1 && && && && && = a_1 \\
&\beta_{i2} && + \beta_{i1}b_1 && + \beta_{i0}b_2 && && && && = a_2 \\
&\beta_{i3} && + \beta_{i2}b_1 && + \beta_{i1}b_2 && + \beta_{i0}b_3 && && && = a_3 \\
(29)\quad &. && . && . && . && . && . && . && . && . \\
&\beta_{in} && + \beta_{in-1}b_1 && + \beta_{in-2}b_2 && + \beta_{in-3}b_3 && + \ldots && + \beta_{i0}b_n && = a_n \\
&\beta_{in+1} && + \beta_{in}b_1 && + \beta_{in-1}b_2 && + \beta_{in-3}b_3 && + \ldots && + \beta_{i1}b_n && = 0 \\
&. && . && . && . && . && . && . && . && . \\
&\beta_{i2n} && + \beta_{i2n-1}b_1 && + \beta_{i2n-2}b_2 && + \beta_{i2n-3}b_3 && + \ldots && + \beta_{in}b_n && = 0
\end{aligned}
$$

which comprise $2n+1$ equations in the $2n+1$ required coefficients of [n/n] viz $\{a_0, a_1, \ldots, a_n; b_1, \ldots, b_n\}$. Baker ([5], Theorem 2.4) has proved that, although a solution to (29) does not necessarily exist for all positive integers n there is an infinite subsequence $\{n_j\}$ for which the Padé approximant $[n_j/n_j]$ exists for any formal power series

with $\beta_{i0} \neq 0$. Further, when the approximant [n/n] exists, it is unique ([5], Theorem 1.1).

We see, by inspection of (29), that in order to compute the coefficients of [n/n] we need the coefficients in the local expansion (26) to order 2n . Even for low values of n , this is likely to become prohibitive quite quickly when dealing with the distribution of an econometric statistic because of the increasing difficulty in extracting higher order coefficients and the complications of the resulting formulae. Moreover, in view of the smoothness of most density functions, in practice we might expect decreasing returns from increasing the order of contact at a particular point past n = 3 or 4. In many cases, n = 2 will be sufficient to provide a highly satisfactory local density approximant.[1]

Multiple point expansions provide an excellent means of enabling us to reduce the order of contact at individual points to within manageable limits while extending the domain over which the final approximant will perform well. Thus, a two point Padé approximant [n/n] might be based on the first n+1 equations of (29) which require local expansion coefficients up to β_{in} and a corresponding set of n equations with expansion coefficients up to order n-1 for a point other than the origin. These equations will then yield an approximant with contact of order n at the origin and n-1 at the second point.

If one of the points of local expansion is infinity, then the equations take on a slightly different form. In this case, rather than (28) we require

$$(30)\quad \{pdf_s(x)Q_n(x) - P_n(x)\}x^{-n} = O(x^{-2n-1})$$

[1] These issues will be taken up in greater depth in later papers.

as $x \to \infty$. We then have the following explicit relations from (25), (27) and (30)

$$\begin{aligned}
&\alpha_0 b_n && && && && = a_n \\
&\alpha_1 b_n && + \alpha_0 b_{n-1} && && && = a_{n-1} \\
&\alpha_2 b_n && + \alpha_1 b_{n-1} && + \alpha_0 b_{n-2} && && = a_{n-2} \\
&\quad . && \quad . && \quad . && \quad . \quad . \quad . \quad . && \quad . \quad . \\
&\alpha_n b_n && + \alpha_{n-1} b_{n-1} && + \alpha_{n-2} b_{n-2} && + \ldots + \alpha_0 && = a_0 \\
&\alpha_{n+1} b_n && + \alpha_n b_{n-1} && + \alpha_{n-1} b_{n-2} && + \ldots + \alpha_1 && = 0 \\
&\quad . && \quad . && \quad . && \quad . \quad . \quad . \quad . && \quad . \quad . \quad . \\
&\alpha_{2n} b_n && + \alpha_{2n-1} b_{n-1} && + \alpha_{2n-2} b_{n-2} && + \ldots + \alpha_n && = 0
\end{aligned}$$

As with (29), this is a system of $2n+1$ equations in the same number of unknown coefficients.

In the general case of expansions about arbitrary points d_i as in (26), the equations which determine the coefficients take the form

$$(32) \quad \Big(\sum_{j=0}^{\infty} \beta_{ij}(x-d_i)^j\Big) Q_n(d_i + (x-d_i)) - P_n(d_i + (x-d_i)) = O((x-d_i)^{2n+1})$$

If we write $y = x - d_i$ and expand $Q_n(d_i + y)$ and $P_n(d_i + y)$ as

$$Q_n(d_i + y) = \sum_{k=0}^{n} b_k (d_i + y)^k = \sum_{k=0}^{n} b_i^{(i)} y^k = Q_n^{(i)}(y)$$

$$P_n(d_i + y) = \sum_{k=0}^{n} a_k (d_i + y)^k = \sum_{k=0}^{n} a_k^{(i)} y^k = P_n^{(i)}(y)$$

we have

$$(33) \quad b^{(i)} = K^{(i)} b \ , \quad a^{(i)} = K^{(i)} a$$

where

$$(34)\quad K^{(i)} = \begin{bmatrix} 1 & d_i & d_i^2 & d_i^3 & \dots & d_i^n \\ 0 & 1 & 2d_i & 3d_i^2 & \dots & \binom{n}{1} d_i^{n-1} \\ 0 & 0 & 1 & 3d_i & \dots & \binom{n}{2} d_i^{n-2} \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot \\ 0 & 0 & 0 & 0 & \dots & 1 \end{bmatrix}, \quad b^{(i)} = \begin{bmatrix} b_0^{(i)} \\ b_1^{(i)} \\ \cdot \\ \cdot \\ \cdot \\ b_n^{(i)} \end{bmatrix}$$

$$b = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{bmatrix} \qquad a^{(i)} = \begin{bmatrix} a_0^{(i)} \\ a_1^{(i)} \\ \vdots \\ a_n^{(i)} \end{bmatrix}, \quad a = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}$$

and (32) becomes

$$(35)\quad \Big(\sum_{j=0}^{\infty} \beta_{ij} y^j\Big) Q_n^{(i)}(y) - P_n^{(i)}(y) = O(y^{2n+1})$$

which has the same form as (28) but in the transformed coefficients.

To work in terms of the original coefficients we can use the transformation matrix $K^{(i)}$ in the case of the vector $a^{(i)}$ as in (33) but in view of the normalization on the vector b we partition $K^{(i)}$ and b as follows

$$(36)\quad K^{(i)} = \left[\begin{array}{c|c} 1 & k^{(i)'} \\ \hline 0 & \\ | & K_{22}^{(i)} \\ | & \end{array}\right], \quad k^{(i)'} = [d_i^1, d_i^2, d_i^3, \dots, d_i^n]$$

$$(37)\quad b' = [b_0, b_*'] = [1, b_*']$$

We then have from (33)

$$(38)\quad b_0^{(i)} = k^{(i)'} b_* \ , \quad b^{(i)} = K_{22}^{(i)} b_*$$

We now define

$$(39)\quad G^{(i)} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ -\beta_{i0} & & & \cdots & 0 \\ -\beta_{i1} & -\beta_{i0} & & \cdots & 0 \\ -\beta_{i2} & -\beta_{i1} & -\beta_{i0} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \\ -\beta_{in-1} & -\beta_{in-2} & -\beta_{in-3} & \cdots & -\beta_{i0} \end{bmatrix}, \quad g^{(i)} = \begin{bmatrix} \beta_{i0} \\ \beta_{i1} \\ \cdot \\ \cdot \\ \vdots \\ \beta_{in} \end{bmatrix}$$

The first $n+1$ equations of (29) can be written as

$$(40)\quad a + G^{(i)} b_* = g^{(i)}$$

or, in the general case of a local expansion about the point d_i (not necessarily the origin)

$$(41)\quad a^{(i)} + G^{(i)} b_*^{(i)} = g^{(i)} + g^{(i)} b_0^{(i)}$$

Transforming back to the original coefficients in the rational fraction we get

$$K^{(i)} a + G^{(i)} K_{22}^{(i)} b_* = g^{(i)} + g^{(i)} k^{(i)'} b_*$$

or

$$(42)\quad K^{(i)} a + [G^{(i)} K_{22}^{(i)} - g^{(i)} k^{(i)'}] b_* = g^{(i)}$$

The system of equations (42) holds for each point of local expansion i.e. for $i = 1, \ldots, I$ in our original notation (see (26)).

Note that as we have constructed (42) the system involves $n+1$ equations. In practice, it may be convenient to use fewer equations at each point, thus reducing the order of contact of the Padé approximant at each point and requiring less analytic information about the expansion coefficients. The procedure allows us to make up for this reduction in the number of equations at each point by increasing the number of points we use in developing the approximant. This process has the additional advantage of improving the global nature of the final approximation.

In the general case we let $K^{(i)}$ be $(m_i+1)\times(n+1)$, $G^{(i)}$ be $(m_i+1)\times m_i$, $K_{22}^{(i)}$ be $m_i \times n$, $g^{(i)}$ be $(m_i+1)\times 1$ and $k^{(i)}$ be $n \times 1$. The complete system of equations which determine the coefficients in the $[n/n]$ Padé approximant are then based on (31) and (42) and take the following general form:

$$
\left[\begin{array}{c|cccc}
K^{(1)} & \multicolumn{4}{c}{G^{(1)}K_{22}^{(1)} - g^{(1)}k^{(1)'}} \\
\hline
 & & & & \\
\hline
K^{(I)} & \multicolumn{4}{c}{G^{(I)}K_{22}^{(I)} - g^{(I)}k^{(I)'}} \\
\hline
 & -\alpha_{r+1} & -\alpha_{r+2} & \cdots & -\alpha_{n+r} \\
 & -\alpha_{r} & -\alpha_{r+1} & \cdots & -\alpha_{n+r+1} \\
0_{r\times(n+1)} & \cdot & \cdot & \cdots & \cdot \\
 & -\alpha_{2} & -\alpha_{3} & \cdots & -\alpha_{n+1} \\
\hline
 & -\alpha_{1} & -\alpha_{2} & \cdots & -\alpha_{n} \\
 & -\alpha_{0} & -\alpha_{1} & \cdots & -\alpha_{n-1} \\
I_{n+1} & 0 & -\alpha_{0} & \cdots & -\alpha_{n-2} \\
 & \cdot & \cdot & \cdots & \cdot \\
 & 0 & 0 & \cdots & -\alpha_{0}
\end{array}\right]
\left[\begin{array}{c}
a_0 \\ a_1 \\ \vdots \\ a_n \\ \hline b_1 \\ b_2 \\ \vdots \\ b_n
\end{array}\right]
=
\left[\begin{array}{c}
g^{(1)} \\ \hline \\ \hline g^{(I)} \\ \hline 0 \\ \vdots \\ 0 \\ \hline -\alpha_0 \\ 0 \\ \vdots \\ 0
\end{array}\right]
\tag{43}
$$

where $m_\infty = n+r+1$. In (43) we need to select I , the m_i and m_∞ in such a way that $\sum_{i=1}^{I} m_i + m_\infty \geq 2n+1$. In the final block of equations in (43) we will often select $r = 0$ so that $m_\infty \leq n+1$. As with the case of the local expansions (26) about the points d_i this will reduce the number of final expansion coefficients that are required to solve for the $[n/n]$ Padé coefficients. In some cases (the application in Section 8 turns out to be such a case) we may have more than enough coefficients, so that $\sum m_i + m_\infty > 2n+1$ and we may neglect some equations of (43) to obtain a solution. When this happens it would seem preferable to neglect those equations which refer to higher order points of contact of the Padé approximant with the local expansions than those which refer to lower order points of contact.

7. MODIFYING THE PADÉ APPROXIMANT

By solving (43) for the Padé coefficients we obtain a preliminary rational fraction [n/n] as in (27). This can be used to construct an approximation to pdf(x) of the form $s(x)[n/n] = s(x)P_n(x)/Q_n(x)$. In some cases, this will turn out to be a perfectly satisfactory approximant. In others, it will need to be modified to produce a good approximation to pdf(x) over a wide interval. This is because there is nothing in the procedure outlined in Section 6 which prevents the occurrence of zeroes in the polynomial $Q_n(x)$ on the real axis. These zeroes induce poles in the approximant and will need to be eliminated if the approximation is to perform well, unless the zeroes appear in remote and irrelevant regions of the distribution. My experience to date suggests that the latter is not usually the case. The normal occurrence is for the procedure in Section 6 to produce a preliminary approximant with either no poles at all or a pair of poles, at least one of which lies within the main body of the distribution.

In addition to unwanted poles, the [n/n] Pade approximant may become zero at a finite number of points on the real axis. Since we will, in general, have $pdf(x) > 0$ for all finite x we will normally wish to eliminate the zeroes of the approximant unless they occur well outside the region of interest in the distribution.

When poles occur, they are typically found in the bridging region between the points of local density expansion used in (43) to construct the Pade approximant. This suggests that an obvious way of helping to remove unwanted poles is to introduce an additional point of local expansion in (43), perhaps at the price of reducing the order of contact at

another point. However, there is no guarantee that this method will eliminate poles and it has the disadvantage of requiring additional information about the distribution to be operational. The procedure we suggest below does not suffer from these disadvantages. It will eliminate the poles and it is sufficiently flexible to allow for additional information about the distribution to be incorporated at the time of modification, if that information is available.

Before we outline the procedure, it may be worth mentioning that the occurrence of unwanted poles in Padé approximants is a long standing problem. The presence of poles in the approximant is one of the reasons why it is difficult to prove general theorems about the convergence of Padé approximants to a given function as the degree of the approximant increases. Many of the general results that are available[1] concern the convergence of subsequences as $n \to \infty$. For an example of non-convergence, we can cite Chui [9] who proves that in the general family of entire functions there exists a function for which the sequence of [n/n] Padé approximants is divergent everywhere in the whole complex plane except at the origin. While such results obstruct the development of a general theory of arbitrarily good approximation by Padé approximants, they do not mean that good approximations cannot in practice be obtained by the Padé procedure in most cases. The modification we give below changes the class of approximants to ensure good global behaviour and eliminate poles at the price of reducing the order of contact at individual points of local expansion.

The procedure we suggest for modifying Padé approximants so that they are well behaved over the whole real axis is based on the following

---

[1] See, for example, Part II of Baker [5].

simple idea. If zeroes of the numerator and denominator polynomials $P_n(x)$ and $Q_n(x)$ occur on the real axis, they will occur in pairs since n is even. We then replace the real roots of the associated quadratic equations by complex conjugate pairs in such a way that we preserve, as far as possible, the known behavior of the function at the points of local expansion. Various degrees of sophistication are possible in the practical application of this method. In fact, as we will demonstrate in the application of Section 8 even crude adjustments which preserve only contact of order one at the points of local expansion seem to work remarkably well. After we have adjusted the coefficients in the Padé approximant so that $P_n(x) > 0$ and $Q_n(x) > 0$ throughout the real axis we simply numerically integrate and rescale so that the area under the curve is unity.

Some of the principles involved in the method outlined in the last paragraph can be illustrated in the case of an approximant with $n = 4$. Let the $[n/n]$ Padé approximant extracted by the procedure of Section 6 with points of local expansion at $x = 0$ and $x^{-1} = 0$ be given by

$$(44) \quad [4/4](x) = \frac{\sum_{i=0}^{4} a_i x^i}{\sum_{i=1}^{4} b_i x^i} = \frac{a_4(x-\gamma_1)(x-\gamma_2)(x-\gamma)(x-\bar{\gamma})}{b_4(x-\delta_1)(x-\delta_2)(x-\delta)(x-\bar{\delta})}$$

where γ_1, γ_2 denote real zeroes of the numerator and δ_1, δ_2 denote real zeroes of the demonator. $(\gamma,\bar{\gamma})$ and $(\delta,\bar{\delta})$ are complex conjugate pairs. We start by rewriting (44) in the form

$$(45) \quad [4/4](x) = \frac{a_4\{x^2 - (\gamma_1+\gamma_2)x + \gamma_1\gamma_2\}(x-\gamma)(x-\bar{\gamma})}{b_4\{x^2 - (\delta_1+\delta_2)x + \delta_1\delta_2\}(x-\delta)(x-\bar{\delta})}$$

We now propose to modify the coefficients of the quadratics in braces so that $[4/4](x) > 0$ for all real x while retaining the same behavior as (45) in the neighborhood of $x = 0$ and $x^{-1} = 0$. We, therefore, define the family of functions

$$(46) \quad [4/4](x;\Theta) = \frac{a_4\{cx^2 + dx + e\}(x-\gamma)(x-\bar{\gamma})}{b_4\{fx^2 + gx + h\}(x-\delta)(x-\bar{\delta})}$$

where $\Theta' = (c,d,e,f,g,h)$ is a vector of real parameters to be chosen. To ensure equivalent local behavior in (45) and (46) we restrict our choice of Θ so that

(i) $c/f = 1$; and

(ii) $e/h = \gamma_1\gamma_2/\delta_1\delta_2$.

Now (i) will ensure that $[4/4](x;\Theta) \to a_4/b_4$ as $x \to \pm\infty$ and (ii) that $[4/4](x;\Theta) \to a_4\gamma_1\gamma_2|\gamma|^2/b_4\delta_1\delta_2|\delta|^2$ as $x \to 0$.

In most cases we find that the zeroes (γ_1, γ_2) and (δ_1, δ_2) occur with the same sign patterns. This is because the zeroes and singularities lie in the intervals between the points (here 0 and $\pm\infty$) of local expansion. If we take the case where both $\gamma_1\gamma_2 > 0$ and $\delta_1\delta_2 > 0$, our task is then to raise f and h from their original values in (45) so that the discriminant $g^2 - 4fh < 0$. This will require proportional changes in c and e so that (i) and (ii) remain valid. Often these automatic changes in c and e will be sufficient to ensure that there are no zeroes in (46). If they are not, some small adjustment in the value of d will normally suffice. There is an added advantage to adjusting the value of d in that simple hand calculations will show what adjustments in this parameter will improve the order of contact of (46) at the points of local expansion while preserving the desired global behavior of $[4/4](x;\Theta) > 0$

for all x. Various other scenarios of parameter changes are possible but those we have illustrated should indicate some of the relevant considerations and the ease with which they may be performed.

The family of rational fractions (46) based on Padé approximants have introduced extra flexibility in the approximating procedure. The idea is essentially to partially reparameterize a first stage Padé approximant so that we can achieve good global behavior by sacrificing some degree of contact at the points of local expansion. But with the new family of approximating rational fractions (46) we have the opportunity to adjust the parameters to take account of any additional information about the distribution that has not already been used in the equations (43) that define the original coefficients; perhaps less precise information based on, for example, Monte Carlo work with the same distribution. Note that we also have the opportunity to modify the Padé approximant by manipulating the new parameters in such a way that the error curve oscillations more closely correspond with the equioscillations that are characteristic of the best approximant within the given class of rational fractions--see Theorem 4 of Section 3.

An obvious alternative procedure for modifying the Padé approximant (45), but which I have not yet tried in application, is to use splines to bridge the intervals in which singularities and zeroes occur. This method may be particularly useful in cases where zeroes and singularities occur together in close proximity. It has the disadvantage that we would no longer be working within the simple family of rational approximants of the form given by (2).

8. AN APPLICATION TO A SIMULTANEOUS EQUATIONS ESTIMATOR

We consider the single structural equation

$$(47)\quad y_1 = \beta y_2 + Z_1\gamma_1 + u$$

where y_1 and y_2 are vectors of T observations on two endogenous variables, Z_1 is a $T \times K_1$ matrix of observations on K_1 exogenous variables and u is a vector of random disturbances. The reduced form equations for y_1 and y_2 are

$$(48)\quad [y_1 \vdots y_2] = [Z_1 \vdots Z_2]\begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix} + [v_1 \vdots v_2]$$

where Z_2 is a $T \times K_2$ matrix of observations of K_2 exogenous variables excluded from (47). We assume that the usual standardizing transformations have been carried out so that (i) $T^{-1}Z'Z = I_K$ where $[Z = Z_1 \vdots Z_2]$, $K = K_1 + K_2$ and (ii) the rows of $[v_1 \vdots v_2]$ are independent and identically distributed normal vectors with zero mean and covariance matrix equal to the identity matrix. We also assume that (47) is identified so that $K_2 \geq 1$.

The two stage least squares estimator 2SLS of β in (47) is given by the ratio $\hat{\beta} = y_2'Ry_1/y_2'Ry_2$ where $R = Z_2Z_2'$. The exact density function of $\hat{\beta}$ is known to be the doubly infinite series [38]

$$(49)\quad \text{pdf}(x) = \frac{1e^{-\frac{\mu^2}{2}(1+\beta^2)}}{\beta\left(\frac{1}{2}, \frac{K_2}{2}\right)(1+x^2)^{(K_2+1)/2}} \sum_{j=0}^{\infty} \frac{\left(\frac{K_2+1}{2}\right)_j}{\left(\frac{K_2}{2}\right)_j j!}\left\{\frac{\mu^2}{2}\frac{(1+\beta x)^2}{1+x^2}\right\}^j$$

$$\cdot {}_1F_1\left(\frac{K_2-1}{2}, j + \frac{K_2}{2}; \frac{\mu^2\beta^2}{2}\right)$$

which depends on the three parameters β , K_2 and $\mu^2 = T\pi'_{22}\pi_{22}$.

The extensive tabulations in Anderson and Sawa [3] show that (49) may be adequately approximated by the asymptotic normal only when μ^2 is very large (the size of μ^2 required for the asymptotic distribution to provide an adequate approximation is itself contingent on the size of β and K_2 . Even for moderate values of β and K_2 , the computations in [3] show that extremely large values of μ^2 (well over 1,000) are required to secure a satisfactory approximation.

As discussed in the Introduction, other approximations to (49) which perform satisfactorily for a range of parameter values are the Edgeworth ([2], [3]) and saddlepoint [21] approximations. But, when μ^2 is small both these approximations become inadequate.

In Figures 2, 3, and 4 we illustrate the inadequacy of these three different methods of approximation in the case where $\beta = 0.6$, $\mu^2 = 4.0$ and $K = 4, 10$. This is rather an extreme case where μ^2 is very low (around the lower limit of μ^2 values found by Anderson *et al.* [4] in their numerical computations of key parameters for actual econometric models). It has been chosen to test the adequacy of the new method of approximation discussed in the earlier sections of this paper specifically in a case where the existing methods break down.

Figures 5-9 detail the approximants obtained at each stage of the procedure outlined in the previous sections of the paper.[1] In the first stage of the procedure we need to select the coefficient function $s(x)$ as in (24) above. A crude choice would be $s(x) = [1 + |x|^{K_2+1}]^{-1}$ since this has the same asymptotic behavior of $|x|^{-(K_2+1)}$, which is in this

[1] We deal specifically with the case $\mu^2 = 4$, $k = 4$, $\beta = 0.6$. Later papers will detail more fully some numerical experience with modified Padé approximants.

case the coefficient function in the tail expansion (16), and since this function is also well behaved elsewhere on the real line. An alternative and better choice is the leading term in the density expansion (49), viz

$$(50)\quad s(x) = \left[B\left(\frac{1}{2}, \frac{K_2}{2}\right)(1+x^2)^{(K_2+1)/2} \right]^{-1}$$

This is, in fact, the pdf of $\hat{\beta}$ under the null hypothesis that $\beta = 0$ and $\pi_{22} = 0$.[1] As mentioned previously in the introduction, leading terms such as (50) in multiple series representations of density functions can usually be derived without much difficulty and will often be available even in cases where an analytic form for the exact density has not been obtained. In the present case, a few elementary manipulations show that $\hat{\beta}$ takes the form of a standard normal variate divided by the square root of a chi square with K_2 degrees of freedom, with the numerator and denominator independent. The statistic $\hat{\beta}$ is, therefore, proportional to a t-variate with K_2 degrees of freedom, leading to a pdf of the form given by (50).

Writing $\text{pdf}(x) = s(x)\text{pdf}_s(x)$ as in (24) we then extract the local expansions (25) and (26) for $\text{pdf}_s(x)$. The expansions we use for the tails $(x^{-1} = 0)$ and the origin $(x = 0)$. The coefficients that appear in (30) and (31) are given by

$$\alpha_0 = \sum_{j=0}^{\infty} \frac{\left(\frac{K+1}{2}\right)_j}{\left(\frac{K}{2}\right)_j j!}\left(\frac{\mu^2}{2}\right)^j W(K,j)$$

[1] Cf. Basmann [8].

$$\alpha_1 = \sum_{j=1}^{\infty} \frac{\left(\frac{K+1}{2}\right)_j 2^{\beta}}{\left(\frac{K}{2}\right)_j (j-1)!}\left(\frac{\mu^2}{2}\right)^j W(K,j)$$

$$\alpha_2 = \sum_{j=1}^{\infty} \frac{\left(\frac{K+1}{2}\right)_j}{\left(\frac{K}{2}\right)_j (j-1)!}\left(\frac{\mu^2}{2}\right)^j [(2j-1)\beta^2 - 1]W(K,j)$$

$$\alpha_3 = \sum_{j=1}^{\infty} \frac{\left(\frac{K+1}{2}\right)_j 2}{\left(\frac{K}{2}\right)_j (j-1)!}\left(\frac{\mu^2}{2}\right)^j \left[\frac{(2j-1)(2j-2)}{3!}\beta^2 - j\right]W(K,j)$$

$$\alpha_4 = \sum_{j=1}^{\infty} \frac{\left(\frac{K+1}{2}\right)_j}{\left(\frac{K}{2}\right)_j (j-1)!}\left(\frac{\mu^2}{2}\right)^j \left[\frac{2(2j-1)(2j-2)(2j-4)}{4!}\beta^4 - j(2j-1)\beta^2 + \frac{(j+1)}{2}\right]W(K,j)$$

and

$$\beta_0 = \sum_{j=0}^{\infty} \frac{\left(\frac{K+1}{2}\right)_j}{\left(\frac{K}{2}\right)_j j!}\left(\frac{\mu^2\beta^2}{2}\right)^j W(K,j)$$

$$\beta_1 = \sum_{j=1}^{\infty} \frac{\left(\frac{K+1}{2}\right)_j}{\left(\frac{K}{2}\right)_j (j-1)!}\left(\frac{\mu^{2j}\beta^{2j-1}}{2^{j-1}}\right)W(K,j)$$

$$\beta_2 = \sum_{j=1}^{\infty} \frac{\left(\frac{K+1}{2}\right)_j}{\left(\frac{K}{2}\right)_j (j-1)!}\left(\frac{\mu^{2j}\beta^{2j-2}}{2^j}\right)(1-\beta^2)W(K,j)$$

$$+ \sum_{j=2}^{\infty} \frac{\left(\frac{K+1}{2}\right)_j 2^{\beta^{-2}}}{\left(\frac{K}{2}\right)_j (j-2)!}\left(\frac{\mu^2\beta^2}{2}\right)^j W(K,j)$$

$$\beta_3 = \sum_{j=1}^{\infty} \frac{\left(\frac{K+1}{2}\right)_j^{(-j)}}{\left(\frac{K}{2}\right)_j (j-1)!} \left(\frac{\mu^{2j}\beta^{2j-1}}{2^{j-1}}\right) W(K,j)$$

$$\beta_4 = \sum_{j=1}^{\infty} \frac{\left(\frac{K+1}{2}\right)_j}{\left(\frac{K}{j}\right)_j (j-1)!} \mu^{2j} \left[\frac{\beta^{2j}(j+1)}{2^{j+1}} - \frac{\beta^{2j-2}}{2^j}\right] W(K,j)$$

$$+ \sum_{j=2}^{\infty} \frac{\left(\frac{K+1}{2}\right)_j}{\left(\frac{K}{2}\right)_j (j-2)!} \left(\frac{\mu^2}{2}\right)^j [\beta^{2j-2} j(j+1) + \tfrac{1}{2}\beta^{2j-4}] W(K,j)$$

where we have dropped the subscript on K_2 for convenience and where

$$W(K,j) = e^{-\frac{1}{2}\mu^2(1+\beta^2)} {}_1F_1\left(\frac{K-1}{2},\ j + \frac{K}{2};\ \frac{\mu^2\beta^2}{2}\right)$$

$$= e^{-\frac{1}{2}\mu^2} {}_1F_1\left(j + \tfrac{1}{2},\ j + \frac{K}{2};\ -\frac{\mu^2\beta^2}{2}\right)$$

by Kummer's transformation [46].

Figure 5 details the local density approximations to pdf(x) based on (25) and (26) with the coefficient function s(x) as in (50). The approximations are good in the locality of the points of expansion, the origin and the tails, but start to deteriorate rapidly as we move out of the immediate vicinity. The right hand tail expansion seems particularly good.

Figure 6 shows the [4/4] Pade approximant to the density (49). This has the form (in the notation of (44))

$$(51) \quad R^{(a)}_{4,4}(x;s) = s(x)[4/4](x) = s(x)\frac{a_4(x-\gamma_1)(x-\gamma_2)(x-\gamma)(x-\bar{\gamma})}{b_4(x-\delta_1)(x-\delta_2)(x-\delta)(x-\bar{\delta})}$$

where

$$a_4 = 4.533619 \qquad b_4 = 1.221628$$

$$\gamma_1 = -1.158240 \qquad \delta_1 = -3.567599$$

$$\gamma_2 = -0.537379 \qquad \delta_2 = -0.485485$$

$$\gamma, \bar{\gamma} = -2.133352 \pm 0.732053i \qquad \delta, \bar{\delta} = 0.310396 \pm 0.613123i$$

As a first step approximant $R^{(a)}_{4,4}(x)$ is rather good, with problem occurring only in the left tail at the singularity $x = \beta_2$ (i.e. $x-\beta = \delta_2 - \beta = -1.085845$ for $\beta = 0.6$) and at the two zeroes $x = \gamma_2, \gamma_1$ (note that the second pole occurs outside the region of immediate interest in the distribution).

The next stage in the procedure is to modify the Padé approximant (51) along the lines suggested in Section 7. The numerical modifications described below represent, in fact, my very first trial efforts in the development of the procedure outlined in Section 7. They should thereform be regarded in the light of a preliminary experiment with a new method. It was indeed the success of this first numerical experiment which stimulated me to explore more fully the potential of the method.

In the denominator of (51) we note that the quadratic $x^2 - (\delta_1 + \delta_2)x + \delta_1\delta_2 = x^2 + 4.05344x + 1.7333$ has discriminant 9.497176. To remove the real zeroes we propose to replace this quadratic by

$$(52) \quad 1.5x^2 + 4x + 3$$

where we have raised the constant and the coefficient of x^2 and simply rounded the coefficient of x . According to the ideas outlined in Section 7, we now need to proportionately adjust the coefficient of x^2 and the constant term in one of the quadratics in the numerator. We select the quadratic $x^2 - (\gamma_1 + \gamma_2)x + \gamma_1\gamma_2 = x + 1.695619x + 0.622414$ which we need to modify in any case to remove the unwanted zeroes of $R^{(a)}_{4,4}(x)$.

Making the proportional adjustments recommended to this quadratic we get $1.5x^2 + 1.695619x + 1.077276$. This gives us the following modified Padé approximant after one change of coefficients

$$(53)\quad R^{(a)}_{4,4}(x) = s(x)\frac{a_4\{1.5x^2+1.695619x+1.077276\}(x-\gamma)(x-\bar{\gamma})}{b_4\{1.5x^2+4x+3\}(x-\delta)(x-\bar{\delta})}$$

This function is graphed in Figure 7 against the exact density. We see that the singularity and zero problems have been eliminated and the performance of the approximation is remarkably good. We note some reduction in the order of contact at the points of local expansion, particularly the origin (or taking into account the change of origin on the graph $x-\beta = -0.6$).

As suggested in Section 7, it is worthwhile to modify at least one of the remaining coefficients to improve the order of contact at the points of local expansion. Note that the success of this procedure can be measured against the original Padé approximant in the relevant localities so we do not need a graph of the exact density to do so. Comparing the Padé and modified Padé approximants in Figures 6 and 7 it is clear that the order of contact of the modified Padé at the origin will be improved if we raise the derivative at this point (i.e., $x = 0$ or $x-\beta = -0.6$). This will be achieved by raising the coefficient of x in the quadratic in braces in the numerator of (53). We make a change in this coefficient from 1.69 to 2.0 (rounding up to the nearest integer), giving the new modified Padé approximant

$$(54)\qquad R^{(b)}_{4,4}(x) = s(x)\frac{a_4\{1.5x^2 + 2.0x + 1.077\}(x-\gamma)(x-\bar{\gamma})}{b_4\{1.5x^2 + 4x + 3\}(x-\delta)(x-\bar{\delta})} .$$

This function is graphed in Figure 8. Even with the rather crude adjustments we have made, (54) is by any standards a good approximation to the true density and is well behaved over the whole real axis. A final adjustment can be made by renormalizing so that the area under (54) is unity. The adjusted curve is displayed in Figure 9.

The error curve for the approximation (54) is shown in Figure 10. The unequal oscillations in this curve indicate that there is scope for further improvement within the same class of approximants. Instead of making additional modifications to the coefficients in (54) to achieve improvements, the general formula (43) in Section 6 was used to find a multiple point Padé approximant of the same degrec as (54) but with interpolations at the points $\beta - 2.0$, $\beta - 1.5$, $\beta - 1.0$, $\beta + 0.5$, $\beta + 1.0$ in addition to the origin and infinity. Thus, we set $I = 6$, $m_1 = 3$ and $m_2 = m_3 = m_4 = m_5 = m_6 = m_\infty = 1$ in (43). The approximation obtained in unmodified form was

$$(55)\qquad R^{(c)}_{4,4}(x) = s(x)\frac{a_4(x-\gamma_1)(x-\bar{\gamma}_1)(x-\gamma_2)(x-\bar{\gamma}_2)}{b_4(x-\delta_1)(x-\bar{\delta}_1)(x-\delta_2)(x-\bar{\delta}_2)}$$

where $a_4 = 7.514124$ $\qquad b_4 = 2.024824$

$\gamma_1, \bar{\gamma}_1 = -0.379459 \pm 1.413745i$ $\qquad \delta_1, \bar{\delta}_1 = 0.385829 \pm 0.905626i$

$\gamma_2, \bar{\gamma}_2 = -0.374932 \pm 0.866617i$ $\qquad \delta_2, \bar{\delta}_2 = 0.139481 \pm 0.700144i$

The approximation (55) has no unwanted zeros or poles and gives three decimal places accuracy over the whole real axis. When graphed against

the exact density, $R^{(c)}_{4,4}(x)$ is in fact indistinguishable from the exact density on the scale of Figures 7-9. The error curve for the approximation $R^{(c)}_{4,4}(x)$ is shown in Figure 11. The scale has been magnified to graph the error function and this has resulted in some lack of smoothness in the function because the exact density was computed by series summation only to five significant figures. The unequal oscillations in the error function displayed here show that there is still scope for further improvement in rational approximants of this degree, in spite of the additional points of interpolation.

9. CONCLUSION

This paper has introduced a new technique of approximating probability density functions. The approximating functions belong to a family of rational fractions and are sufficiently flexible to be capable of producing good approximants to a very wide class of density functions. The theory developed in Section 3 indicates that this family of rational fractions contains approximants which are best in a well defined sense and which will perform well in reproducing the form of the exact density functions over the entire real axis. The practical procedure for finding good approximants in this family that is discussed in the paper is based on the use of multiple-point Padé approximants to construct global approximations from purely local information about the density. These multiple-point Padé approximants are then modified to ensure that they have good global behavior and to incorporate any additional information that may be available concerning the density. The application in Section 8 illustrates that the procedure can produce exceptionally good approximations even in cases where existing methods break down. In this application, the multiple-

point Padé approximants themselves show evidence that there is considerable room for further improvement within the family of rational functions of the same degree. In practical application, the extent to which further improvement is possible or desirable will depend on the information that is available about the exact density and the level of accuracy required. Practical application of the methods discussed in this article will also require general programmable formulae for the coefficients in the rational approximants in terms of relevant model and data set. Some applications will also require direct approximants to the df and to tail area probabilities rather than the pdf . These are issues to be investigated in later work.

APPENDIX

We start with some preliminaries to the proof of existence in Theorem 1. We need to show that there exists a set of parameters γ^* and a corresponding rational fraction belonging to (2) for which

$$(A1) \quad \|R_{m,n}(x;s,\gamma^*) - pdf(x)\| = \inf_{\gamma\in\Gamma} \|R_{m,n}(x;s,\gamma) - pdf(x)\| = \rho \ , \quad \text{say,}$$

Our approach is based on the work of Achieser [1] and Rice [37]. Unfortunately, neither of these standard works give results which can be directly used here. Achieser's proof of existence in [1] on pages 53-54 is somewhat defective because he does not allow for the possibility that γ^* may not lie in Γ and may not give rise immediately to a rational fraction of the form (2). Rice's proof is given explicitly for approximation on the finite interval $[0,1]$ and for the class (2) in which $s(x) = 1$. We therefore show the development of the existence proof in full, following in the main the approach taken by Rice [37] in Chapter 3.8.

From (4) we have $0 \leq \rho < \infty$ and we can find a sequence of rational fractions $\{R_{m,n}(x;s,\gamma^{(j)})\}$ for which

$$(A2) \quad \rho_j = \|R_{m,n}(x;s,\gamma^{(j)}) - pdf(x)\|$$

and

$$(A3) \quad \lim_{j\to\infty} \rho_j = \rho \ .$$

It remains to prove that the parameter sequence $\{\gamma^{(j)}\}$ has a convergent subsequence which converges to a set of finite parameters. If we call the latter γ^* then it will follow from (A3) that γ^* satisfies (A1). As discussed by Rice ([37], pp. 26-27), the crucial part of the proof

of existence is to demonstrate that the parameters lie in a compact set. First, we show that we may restrict our attention to bounded subsets of

DEFINITION (Condition E of Rice: [ ], p. 27). *The approximating function* $R_{m,n}(x;s,\ )$ *is said to satisfy Condition E for the norm* $\|\ \|$ *if, given* $M \to \infty$, *there is an* $N \to \infty$ *such that*

$$\|R_{m,n}(x;s,\gamma)\| \leq M$$

implies that

$$\max_i |\gamma_i| \leq N$$

where $\gamma = (\gamma_i)$.

In view of (A2) and (A3), there is an integer j_0 for which

(A4) $$\|R_{m,n}(x;s,\gamma^{(j)}) - \mathrm{pdf}(x)\| \leq \rho+1$$

for all $j > j_0$. Moreover, from Assumption 1 it follows that there exists $K > 0$ for which $\|\mathrm{pdf}(x)\| \leq K$ and, hence, using (A4) we have the inequality

(A5) $$\|R_{m,n}(x;s,\gamma^{(j)})\| \leq K+\rho+1 .$$

We now verify that Condition E holds for the approximating function $R_{m,n}(x;s,\gamma)$.

LEMMA 1 *The rational fraction* $R_{m,n}(x;s,\gamma)$ *defined by* (2) *satisfies Condition E for the uniform norm* (3)

Proof

We consider the set

$$(A6) \quad \{\gamma : \|R_{m,n}(x;s,\gamma)\| \leq M,\ M > 0\}$$

Since $\|R_{m,n}(x;s,\gamma)\| \leq M$ implies that, for a given number $L > 0$,

$$\max_{x\in[-L,L]} |R_{m,n}(x;s,\gamma)| \leq M$$

it follows that (A6) lies in the set

$$(A7) \quad \{\gamma : \max_{x\in[-L,L]} |R_{m,n}(x;s,\gamma)| \leq M\}$$

Now for $x \in [-L,L]$ and taking $L > 1$, we have

$$|R_{m,n}(x;s,\gamma)| \geq \frac{\max_{x\in[-L,L]} |s(x)|}{\max_{x\in[-L,L]} |Q_n(x)|} |P_m(x)|$$

$$(A8) \qquad \geq \frac{s_L (L^2-1)^{1/2}}{(L^{2(n+1)}-1)^{1/2}} |P_m(x)| \ .$$

where

$$s_L = \max_{x\in[-L,L]} |s(x)| > 0 \ .$$

Thus, when γ lies in the set (A7), we have

$$\frac{M(L^{2(n+1)}-1)^{1/2}}{s_L(L^2-1)^{1/2}} \geq \max_{x\in[-L,L]} |P_m(x)| = \max_{x\in[-L,L]} \left| \sum_{i=0}^{m} a_i x^i \right|$$

and the polynomial $\sum_{i=0}^{m} a_i x^i$ is bounded uniformly on the interval $[-L,L]$. It follows that the coefficient parameters a_i are also bounded. Moreover, $\sum_{i=0}^{n} b_i^2 = 1$ by definition, so that the parameter set (A7) is bounded. By implication, the same is true for the set (A6) and, thus, $R_{m,n}(x;s,\gamma)$ satisfies Condition E for the uniform norm. □

Proof of Theorem 1 (Existence)

Lemma 1 shows that we can confine our attention to bounded subsets of Γ in searching for a best approximant. Hence, for all $j \geq j_0$ the sequence $\{\gamma^{(j)}\}$ introduced earlier lies in a bounded subset of Γ. We may, therefore, select a subsequence that converges to the vector γ^*. If we reindex the subsequence we can write for the individual components of γ, $\lim_{j\to\infty} a_i^{(j)} = a_i^*$ and $\lim_{j\to\infty} b_k^{(j)} = b_k^*$ for $i = 0, 1, \ldots, m$ and $k = 0, 1, \ldots, n$.

Now it is important to note that since Γ is not closed γ^* may or may not lie in Γ. If $\gamma^* \in \Gamma$ then $R_{m,n}(x;s,\gamma^*)$ is a rational function of the form defined in (2); and, in view of (4), this is a best uniform approximant of $pdf(x)$. But, if $\gamma^* \notin \Gamma$ then $R_{m,n}(x;s,\gamma^*)$ is the limit of a sequence of rational functions and is not necessarily a rational function itself. In fact, it may not even be continuous (we give an example later in this section). However, the limit function $R_{n,n}(x;s,\gamma^*)$ will differ from a rational function, $R'_{n,n}(x;s,\gamma^*)$ say, only at a finite number of points. And, in fact, this rational function $R'_{n,n}(x;s,\gamma^*)$ is a best uniform approximant to $pdf(x)$ in the class defined by (2).

The problem discussed in the last paragraph arises because although the denominator polynomial $Q_n(x) = Q_n(x;\gamma) > 0$ for $\gamma \in \Gamma$, this no longer necessarily holds when γ does not lie in Γ. Since γ^* is a subsequential limit of elements of Γ it follows that, in the limit,

$Q_n^*(x) = Q_n(x;\gamma^*)$ can have at most n zeroes. If we let $P_n^*(x) = P_n(x;\gamma^*)$ be the limit of the numerator polynomial as $\gamma^{(j)} \to \gamma^*$ it follows that there are at most n points where $R_{m,n}(x;s,\gamma^*) = s(x)P_m^*(x)/Q_n^*(x)$ is undefined. At all other points we must have $R_{m,n}(x;s,\gamma^{(j)}) \to R_{m,n}(x;s,\gamma^*)$ as $j \to \infty$. Moreover, since $R_{m,n}(x;s,\gamma^*) = \text{pdf}(x) + \{R_{m,n}(x;s,\gamma^{(j)}) - \text{pdf}(x)\} + \{R_{m,n}(x;s,\gamma^*) - R_{m,n}(x;s,\gamma^{(j)})\}$ it follows from (A4) that for $j \geq j_0$

$$|R_{m,n}(x;s,\gamma^*)| \leq K + \rho + 1 + |R_{m,n}(x;s,\gamma^*) - R_{m,n}(x;s,\gamma^{(j)})|$$

and allowing $j \to \infty$ we deduce that

$$\text{(A9)} \quad |R'_{m,n}(x;s,\gamma^*)| \leq K + \rho + 1$$

Hence, for all x other than zeroes of $Q_n^*(x)$, we have the inequality

$$\text{(A10)} \quad s(x)|P_m^*(x)| \leq (K+\rho+1)Q_n^*(x) .$$

By continuity, (A10) holds also when $Q_n^*(x) = 0$. Thus any real zero of $Q_n^*(x)$ is also a zero of $P_m^*(x)$, since $s(x) > 0$ for all finite x. We, therefore, eliminate by cancellation each linear factor of $Q_n^*(x)$ corresponding to a real root of $Q_n^*(x) = 0$. We call the resulting rational fraction $R'_{n,n}(x;s,\gamma^*)$ and note that for all values of x other than zeroes of $Q_n^*(x)$

$$R'_{m,n}(x;s,\gamma^*) = R_{m,n}(x;s,\gamma^*)$$

while at the zeroes $\{x_k : k = 1, \ldots, k_0\}$ of $Q_n^*(x)$

$$R'_{m,n}(x_k;\, s,\, \gamma^*) = \lim_{x \to x_k} R_{m,n}(x;s,\gamma^*)$$

Finally, we note that

$$\|R_{m,n}(x;s,\gamma^*) - pdf(x)\| = \sup_{x\in(-\infty,\infty)} |R_{m,n}(x;s,\gamma^*) - pdf(x)|$$

$$= \max\{ \sup_{\substack{x\in(-\infty,\infty) \\ x\neq x_k}} |R_{m,n}(x;s,\gamma^*) - pdf(x)|,\ |R_{m,n}(x_k;s,\gamma^*) - pdf(x_k)|,$$

$$i = 1, \ldots, m\}$$

$$\geq \sup_{x\in(-\infty,\infty)} |R'_{n,n}(x;s,\gamma^*) - pdf(x)| = \|R_{n,n}(x;s,\gamma^*) - pdf(x)\| .$$

The rational function $R'_{m,n}(x;s,\gamma^*)$ is, therefore, a best uniform approximation of $pdf(x)$. □

An Example of Degeneracy

To illustrate the problem of degeneracy in the rational approximant that arises in the proof of this theorem because Γ is not closed, we consider the following density function of the Pareto distribution

$$pdf(x) = \frac{ak^a}{x^{a+1}} , \quad a > 0 , \quad x \geq k > 0 .$$

We consider the case in which $a = 2$ and a class of rational approximants of the form (2) is being used with $s(x) = 1/x$ $(x \geq k)$ and $n = 4$. Now consider the sequence of approximants defined by

$$R_{4,4}(x;s,\gamma^{(j)}) = \left(\frac{2k^2}{x}\right) \frac{\left(\frac{1}{1+k^4} - \frac{1}{2j^2}\right)^{1/2} x^2 - \left(\frac{k^4}{1+k^4} - \frac{1}{2j^2}\right)^{1/2}}{\left(\frac{1}{1+k^4} - \frac{1}{2j^2}\right)^{1/2} x^4 - \left(\frac{k^4}{1+k^4} - \frac{1}{2j^2}\right)^{1/2} x^2 + \frac{1}{j}}$$

$$\gamma^{(j)\prime} = \left[-\left(\frac{k^4}{1+k^4} - \frac{1}{2j^2}\right)^{1/2}, 0, \left(\frac{1}{1+k^4} - \frac{1}{2j^2}\right)^{1/2}, 0, 0; \frac{1}{j}, 0, \right.$$
$$\left. -\left(\frac{k^4}{1+k^4} - \frac{1}{2j^2}\right)^{1/2}, 0, \left(\frac{1}{1+k^4} - \frac{1}{2j^2}\right)^{1/2} \right]$$

and, as $j \to \infty$,

$$\gamma^{(j)\prime} \to \gamma^{*\prime} = \left[-\left(\frac{k^4}{1+k^4}\right)^{1/2}, 0, \left(\frac{1}{1+k^4}\right)^{1/2}, 0, 0; 0, 0, \right.$$
$$\left. -\left(\frac{k^4}{1+k^4}\right)^{1/2}, 0, \left(\frac{1}{1+k^4}\right)^{1/2} \right]$$

The limiting function is then

$$R_{4,4}(x;s,\gamma^*) = \begin{cases} \dfrac{2k^2}{x^3}, & x > k \\ 0, & x = k \end{cases}$$

which is not a rational function nor is it continuous on the interval $[k,\infty)$. However

$$R'_{4,4}(x;s,\gamma^*) = \frac{2k^2}{x^3}, \quad x \geq k$$

is rational, continuous and is clearly the best uniform approximant to pdf(x) on $[k,\infty)$. □

The following version of the Weierstrass Theorem on the uniform approximation of periodic functions by trigonometric sums is useful in the proof of Theorem 2.

LEMMA 2. *The system of trigonometric sums*

$$(A11)\quad a_0 + \sum_{k=1}^{n} (a_k \cos k\theta + b_k \sin k\theta)$$

is dense in the space of continuous periodic functions with period 2π .

Proof (Meinardus [25], p. 8). □

PROOF OF THEOREM 2

Since $R'_{nn}(x;s,\gamma^*)$ is the best uniform approximant to $pdf(x)$ we have the inequality

$$(A12)\quad E(n,s) \leq \|pdf(x) - Q_n(x)\|$$

where

$$(A13)\quad Q_n(x) = s(x)\frac{P_n(x)}{(1+x^2)^{n/2}} \quad (n = \text{even integer})$$

for every choice of polynomial $P_n(x)$ and for every value of n (even). We now substitute variables according to the transformation

$$\theta = 2 \tan^{-1} x$$

which takes the interval $-\infty \leq x \leq \infty$ into $-\pi \leq \theta \leq \pi$. We have

$$Q_n(\tan(\theta/2)) = s(\tan(\theta/2))[a_0\cos^n(\theta/2) + a_1\cos^{n-1}(\theta/2)\sin(\theta/2) + \ldots + a_n\sin^n(\theta/2)]$$

and using the relations

$$\sin\theta = 2\sin(\theta/2)\cos(\theta/2)$$

$$\sin^{2r}(\theta/2) = 2^{-2r}\left\{\sum_{j=0}^{r-1}(-1)^{r-j}2\binom{2r}{j}\cos((r-j)\theta) + \binom{2r}{r-1}\right\}$$

$$\cos^{2r}(\theta/2) = 2^{-2r}\left\{\sum_{j=0}^{r-1}2\binom{2r}{j}\cos((r-j)\theta) + \binom{2r}{r-1}\right\}$$

we deduce that (A13) can be represented as

$$Q_n(\tan(\theta/2)) = s(\tan(\theta/2))S_{n/2}(\theta) \ , \quad -\pi \leq \theta \leq \pi$$

where $S_{n/2}(\theta)$ is a trigonometric sum of the form (A11).

We deal first with the case in which $pdf(x)/s(x)$ tends to a finite limit as $x \to \pm\infty$. Setting $pdf^*(\theta) = pdf(\tan(\theta/2))$ and $s^*(\theta) = s(\tan(\theta/2))$ for $-\pi \leq \theta \leq \pi$, we have

$$\|pdf^*(\theta) - s^*(\theta)S_{n/2}(\theta)\| \leq \|s^*(\theta)\| \left\|\frac{pdf^*(\theta)}{s^*(\theta)} - S_{n/2}(\theta)\right\|$$

where the uniform norm is now taken over the interval $[-\pi, \pi]$. Since $pdf^*(\theta)/s^*(\theta)$ is a continuous function over $-\pi \leq \theta \leq \pi$ and by Lemma 2 can be uniformly well approximated by trigonometric sums of the type $S_{n/2}(\theta)$, it follows that there exists an n_0 and a trigonometric function $S_{n_0/2}(\theta)$ for which

$$\left\|\frac{pdf^*(\theta)}{s^*(\theta)} - S_{n_0/2}(\theta)\right\| < \frac{\varepsilon}{C}$$

where $C = \|s^*(\theta)\|$ and $\varepsilon > 0$ is arbitrarily small. Hence, $\|pdf^*(\theta) - s^*(\theta)S_{n_0/2}(\theta)\| < \varepsilon$.

Now consider the case where $pdf(x)/s(x)$ becomes infinitely large as $x \to \pm\infty$. We split the coefficient function into two factors as

$s(x) = s_1(x)s_2(x)$, where both $s_1(x)$ and $s_2(x)$ are positive functions which $\to 0$ as $x \to \pm\infty$ and where $pdf(x)/s_1(x)$ tends to a finite limit as $x \to \pm\infty$. (For example, $s_1(x)$ could be a positive constant multiple of $pdf(x)$.) Then, there exists an $M > 0$ for which $\|pdf(x)/s(x)\| = M < \infty$. We define $s_1^*(\theta) = s_1(\tan(\theta/2))$ and $s_2^*(\theta) = s_2(\tan(\theta/2))$ for $-\pi \leq \theta \leq \pi$. Given $\varepsilon > 0$ arbitrarily small, since $s_1^*(\theta)$, $s_2^*(\theta)$ both $\to 0$ as $|\theta| \to \pi$, we can find $\theta_1, \theta_2 > 0$ such that

$$s_1^*(\theta) \leq \varepsilon/4M \quad \text{for all} \quad \theta \quad \text{in} \quad [-\pi, -\theta_1] \quad \text{and} \quad [\theta_2, \pi]$$

$$s_2^*(-\theta_1) = s_2^*(\theta_2) = \varepsilon$$

and

$$s_2^*(\theta) \leq \varepsilon \quad \text{for all} \quad \theta \quad \text{in} \quad [-\pi, -\theta_1] \quad \text{and} \quad [\theta_2, \pi]$$

We let Θ_{12} denote the interval $(-\theta_1, \theta_2)$ and Θ_{12}^c its complement in $[-\pi, \pi]$. We define the continuous function

$$s_2^{**}(\theta) = \begin{cases} s_2^*(\theta) , & \theta \in \Theta_{12} \\ \varepsilon , & \theta \in \Theta_{12}^c \end{cases}$$

Now, since $pdf^*(\theta)/s_1^*(\theta)s_2^{**}(\theta)$ is continuous on $[-\pi, \pi]$ there exists a trigonometric sum of the form $S_{n/2}(\theta)$ and an n_0 for which

$$\sup_{\theta \in [-\pi, \pi]} \left| \frac{pdf^*(\theta)}{s_1^*(\theta)s_2^{**}(\theta)} - S_{n_0/2}(\theta) \right| < \frac{\varepsilon}{2D}$$

where

$$D = \sup_{\theta \in [-\pi, \pi]} s_1^*(\theta)s_2^{**}(\theta) .$$

It follows that

$$\sup_{\theta\in\Theta_{12}} |pdf^*(\theta) - s_1^*(\theta)s_2^*(\theta)S_{n_0/2}(\theta)| = \sup_{\theta\in\Theta_{12}} |pdf^*(\theta) - s_1^*(\theta)s_2^{**}(\theta)S_{n_0/2}(\theta)|$$

$$< D\left(\frac{\varepsilon}{2D}\right) = \frac{\varepsilon}{2}$$

and

$$\sup_{\theta\in\Theta_{12}^c} |pdf^*(\theta) - s_1^*(\theta)s_2^*(\theta)S_{n_0/2}(\theta)| \leq \sup_{\theta\in\Theta_{12}^c} |pdf^*(\theta) - s_1^*(\theta)s_2^{**}(\theta)S_{n_0/2}(\theta)|$$

$$+ \sup_{\theta\in\Theta_{12}^c} \{s_1^*(\theta)|S_{n_0/2}(\theta)||s_2^{**}(\theta) - s_2^*(\theta)|\}$$

$$< D\left(\frac{\varepsilon}{2D}\right) + \left(\frac{\varepsilon}{4M}\right)\left(\frac{2M}{\varepsilon}\right)(\varepsilon) = \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

since

$$\sup_{\theta\in\Theta_{12}^c} |S_{n_0/2}(\theta)| \leq \sup_{\theta\in\Theta_{12}^c} \left|\frac{pdf^*(\theta)}{s_1^*(\theta)s_2^{**}(\theta)} - S_{n_0/2}(\theta)\right| + \sup_{\theta\in\Theta_{12}^c} \left|\frac{pdf^*(\theta)}{s_1^*(\theta)s_2^{**}(\theta)}\right|$$

$$\leq 2 \sup_{\theta\in\Theta_{12}^c} \left|\frac{pdf^*(\theta)}{s_1^*(\theta)s_2^{**}(\theta)}\right| \leq \frac{2M}{\varepsilon} .$$

Hence, given $\varepsilon > 0$ arbitrarily small there exists an n_0 and a trigonometric sum $S_{n_0/2}(\theta)$ for which $\|pdf^*(\theta) - s^*(\theta)S_{n_0/2}(\theta)\| < \varepsilon$. It follows from (A12) that $E(n,s) < \varepsilon$ for all $n \geq n_0$.

□

PROOF OF THEOREM 3

de la Vallée-Poussin first proved this theorem for polynomial and trigonometric function approximants. Achieser's proof in [1] on pp. 52-53 for rational fractions holds also for our class of rational fractions (2).

PROOF OF THEOREM 4 (outline)

The proofs of Part B (sufficiency), Part A(i) (necessity in the numerator most degenerate case) and Part A(iii) follow as in Achieser [1], pp. 55-56. The difference with the classical theory in Achieser arises in Part A(ii). In this case, when the denominator is most degenerate we cannot use the construction that appears in the classical proof and, as the example in the body of the paper shows, the classical result on the necessary number of alternations does not in general apply. A full proof of the necessary condition given here in Part A(ii) together with a discussion of special cases in which the classical result still applies will appear elsewhere (Phillips [36]) and can be obtained from the author on request. □

PROOF OF THEOREM 5.

Achieser [1], pp. 56-57. □

PROOF OF THEOREM 6

The first statement follows by dominated convergence from the existence of the $(M-1)^{th}$ absolute moment. The behavior of the derivatives at $\pm\infty$ follows from the representation

$$cf^{(n)}(s) = \int_{-\infty}^{\infty} e^{isx}(ix)^n pdf(x)dx$$

and, since $(ix)^n pdf(x)$ is absolutely integrable on $(-\infty,\infty)$, the Riemann-Lebesgue Lemma ensures that $cf^{(n)}(s) \to 0$ as $s \to \pm\infty$ for each $n = 0, 1, \ldots, M-1$. □

PROOF OF THEOREM 7

By integration by parts, following the argument in Erdéyli [15], p. 47. □

PROOF OF THEOREM 8

By the theory of Fourier transforms of generalized functions and their asymptotic expansions as developed in Lighthill [23]. Full details of the derivations are given in [35], which is available from the author on request. □

PROOF OF THEOREM 9

Since $s^N cf(s)$ is absolutely integrable we expand the exponential e^{-isx} in the inversion formula

$$pdf(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-isx} cf(s)ds$$

about the point $x = d_i$ and integrate term by term giving

$$\frac{1}{2\pi j!}\sum_{j=0}^{N-1}\int_{-\infty}^{\infty}(-is)^j e^{-isd_i} cf(s)ds(x-d_i)^j + \frac{1}{2\pi N!}\int_{-\infty}^{\infty}(-is)^N e^{-isd_i} cf(s)ds(x-d_i)$$

where d_i lies on the line segment connecting x and d_i. □

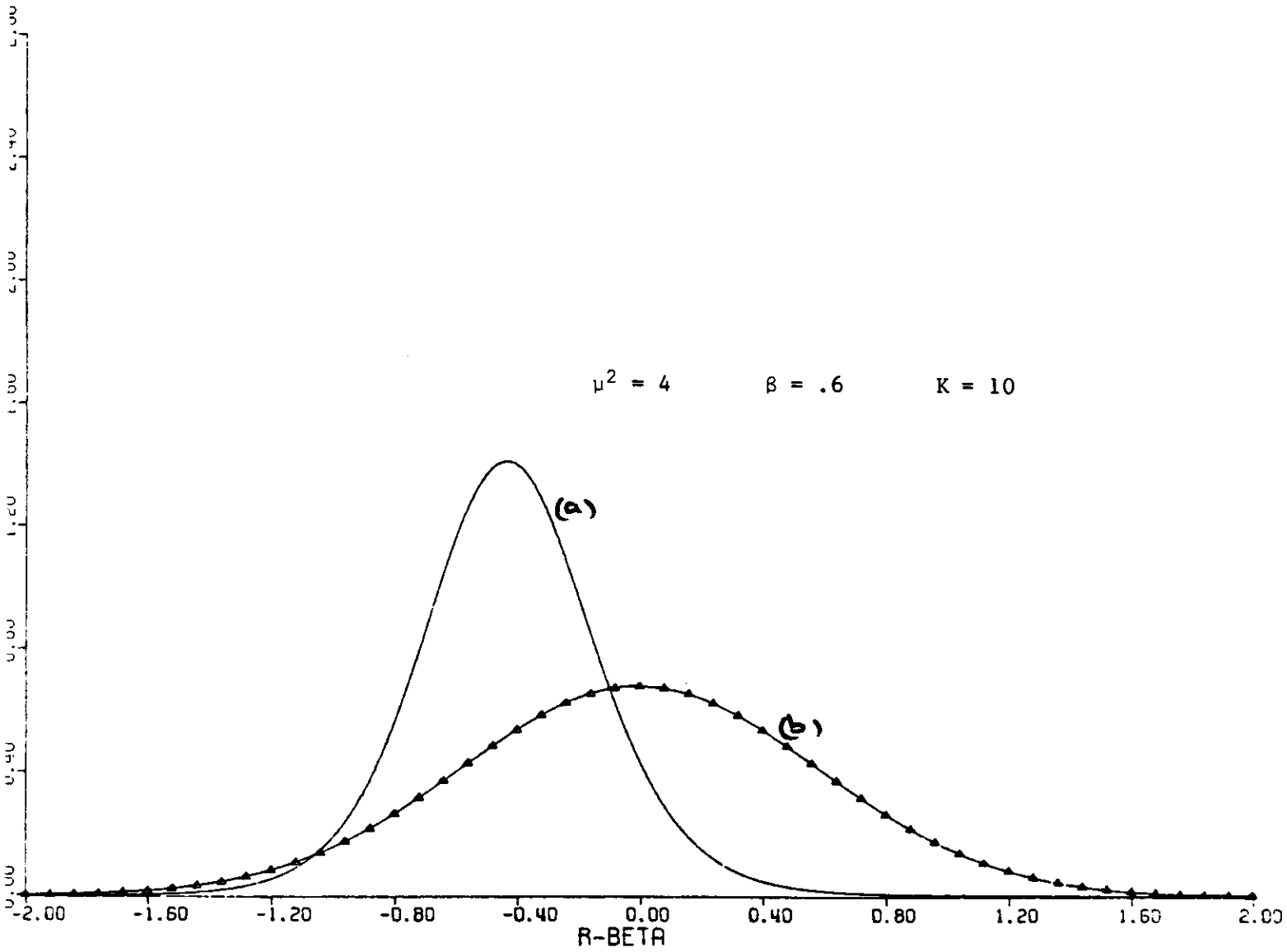

FIGURE 2B Asymptotic Normal Approximation

(a) exact density
(b) asymptotic normal

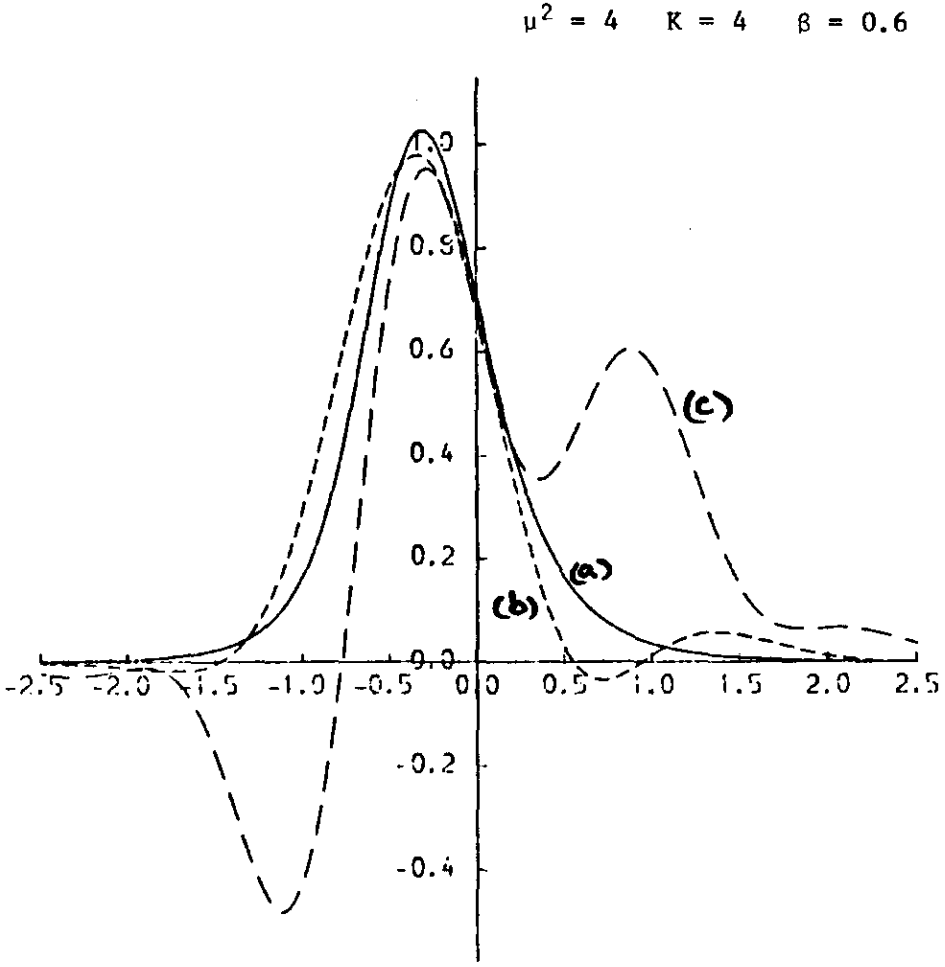

FIGURE 3A Edgeworth Approximation

(a) exact density
(b) $O(T^{-1/2})$
(c) $O(T^{-1})$

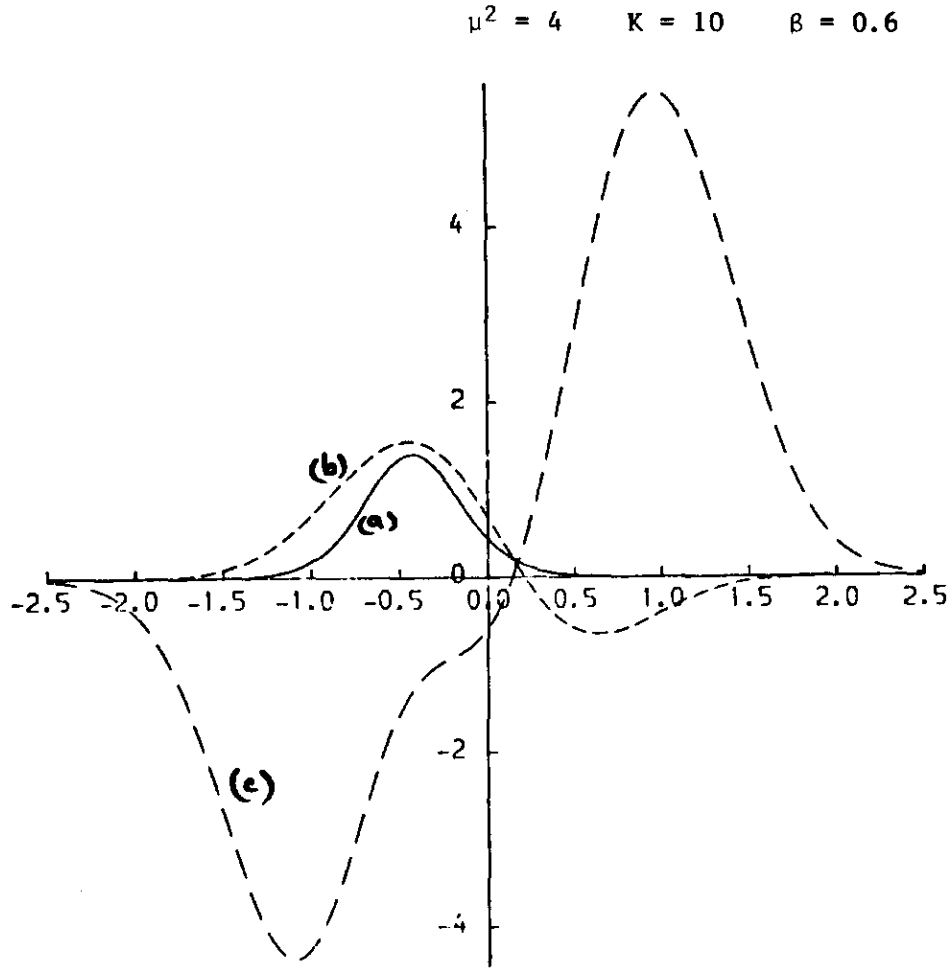

FIGURE 3B Edgeworth Approximation

(a) exact
(b) $O(T^{-1/2})$
(c) $O(T^{-1})$

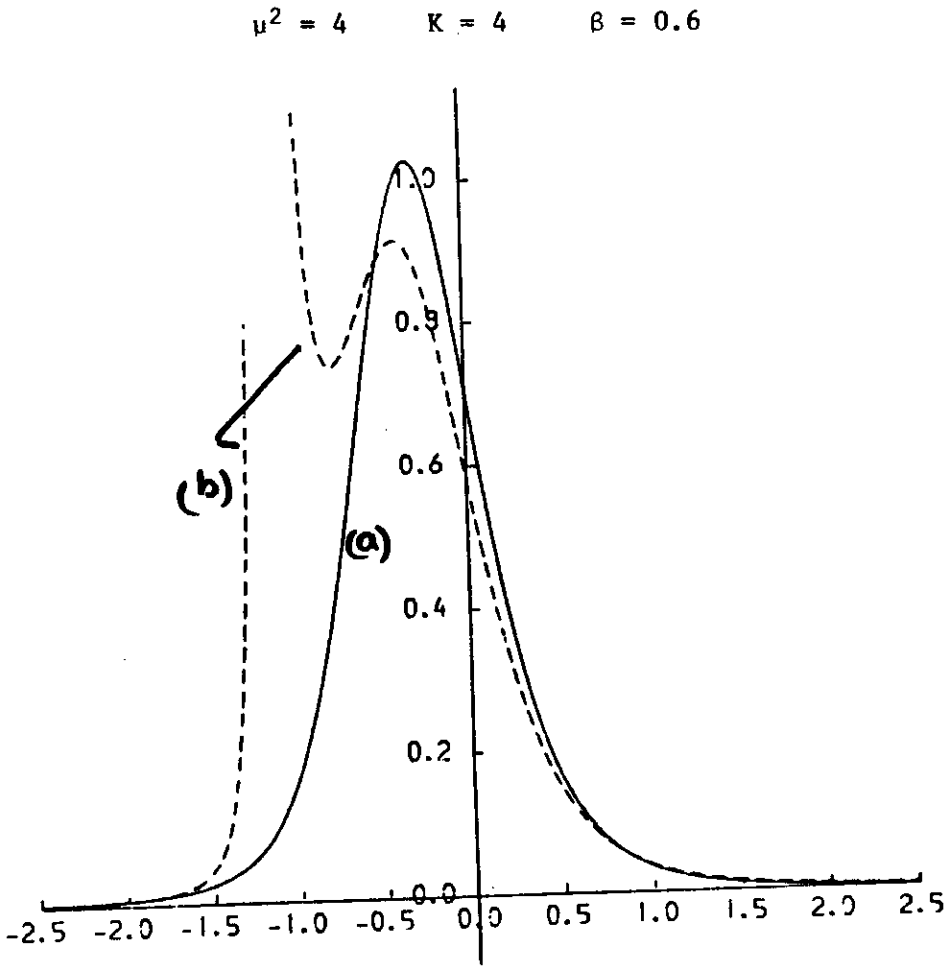

FIGURE 4A Saddlepoint Approximation

(a) exact density
(b) saddlepoint

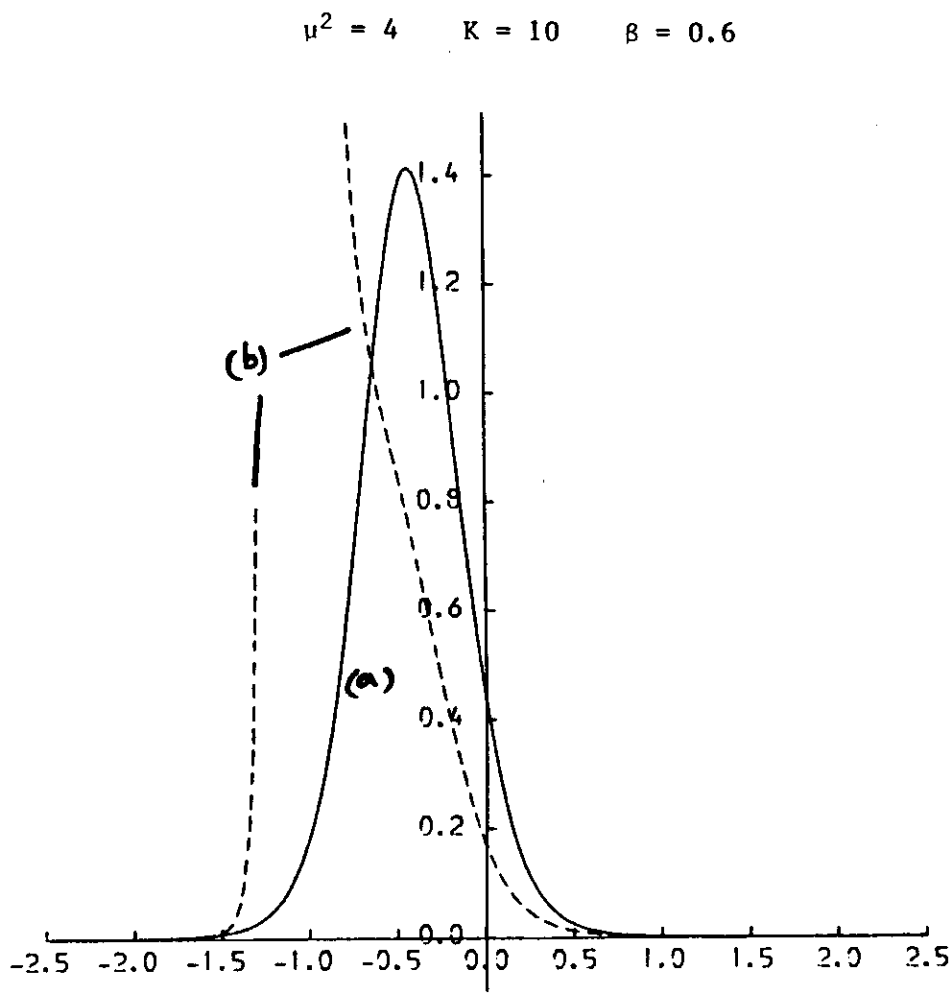

FIGURE 4B Saddlepoint Approximation

(a) exact density
(b) saddlepoint

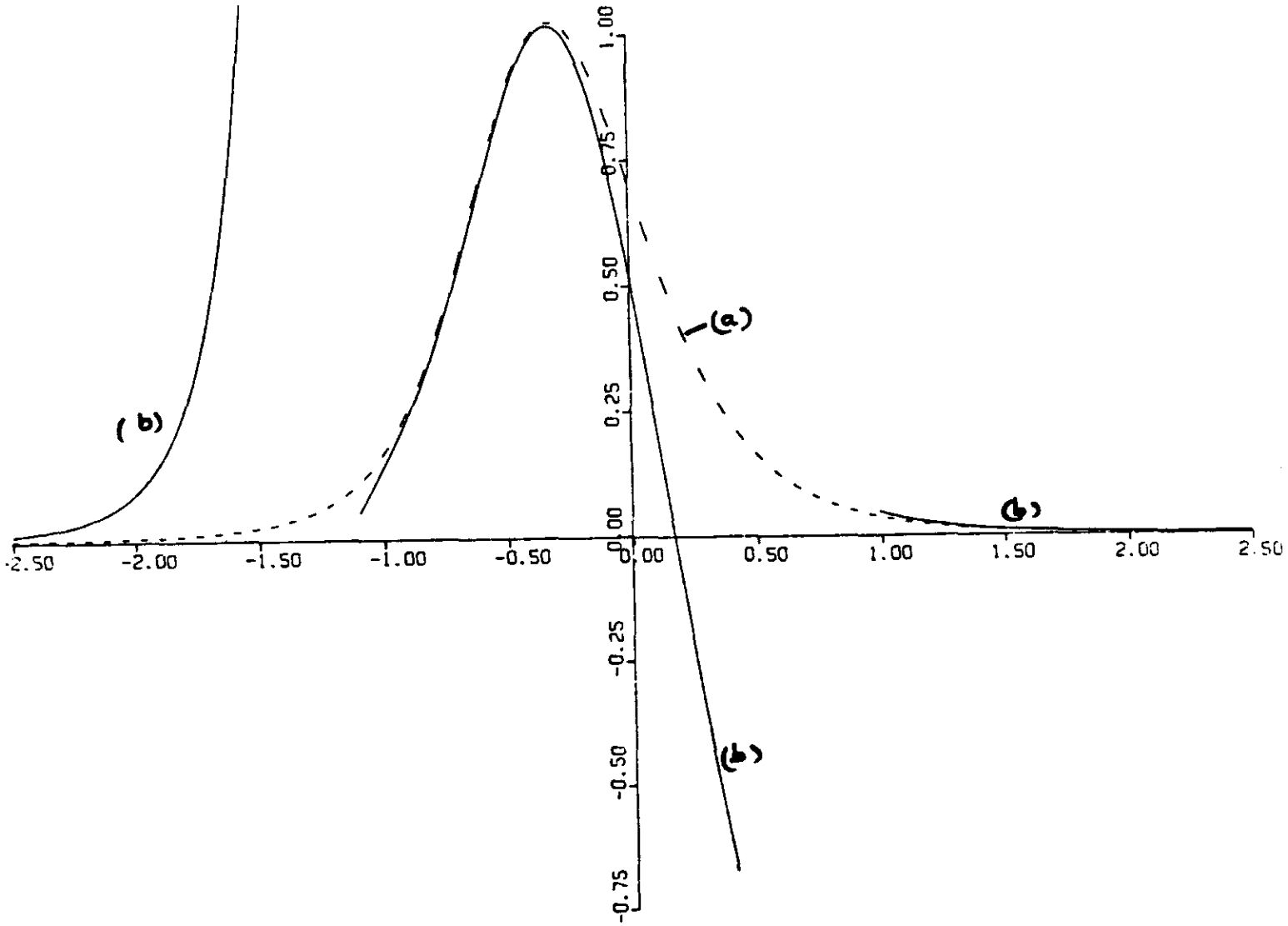

FIGURE 5 Local density approximations at the origin and in the tails to pdf(x)

(a) exact density

(b) local approximations

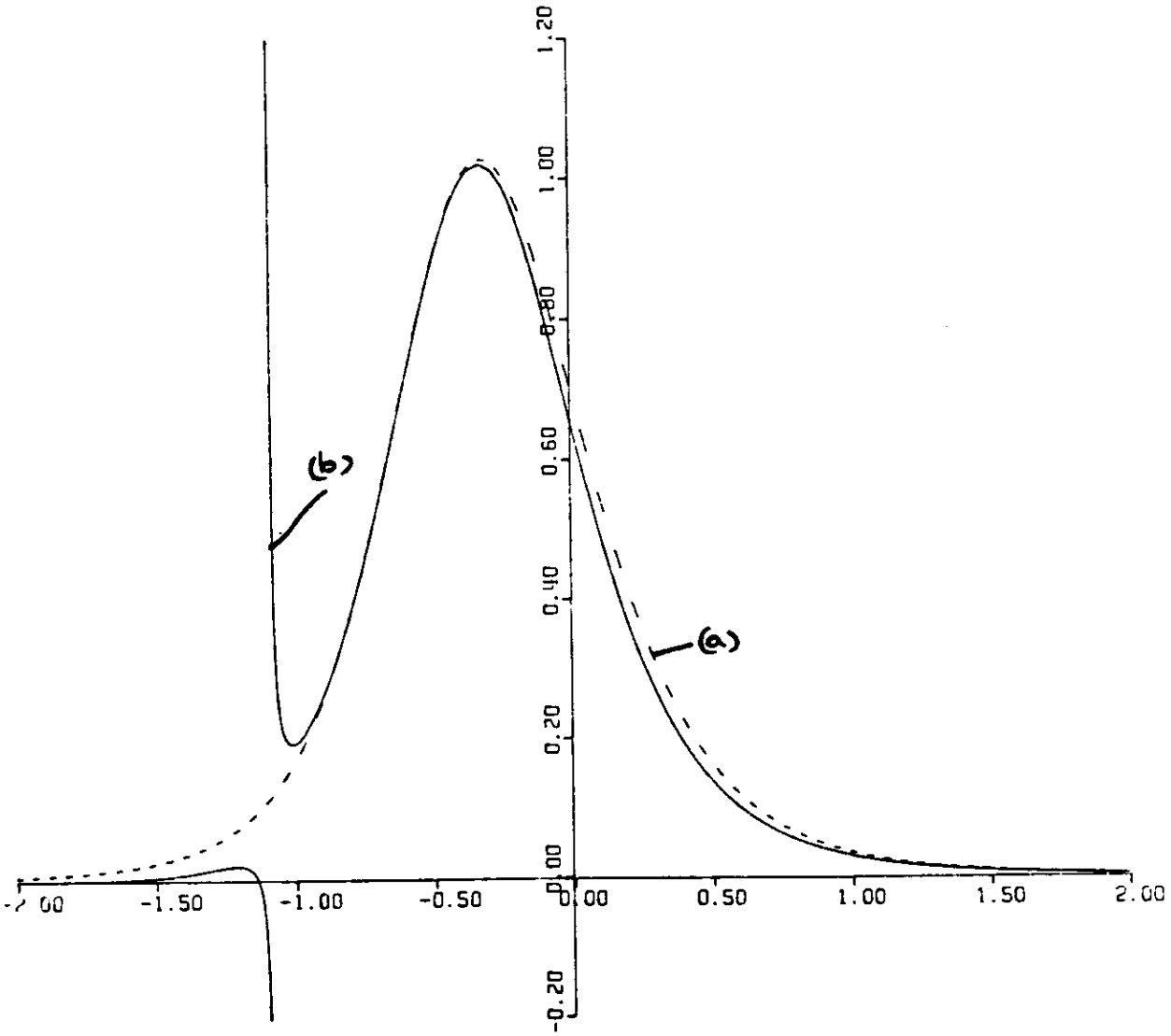

FIGURE 6 First step ⌈4/4⌉ Pade approximant to pdf(x)

(a) exact density

(b) Padé approximant

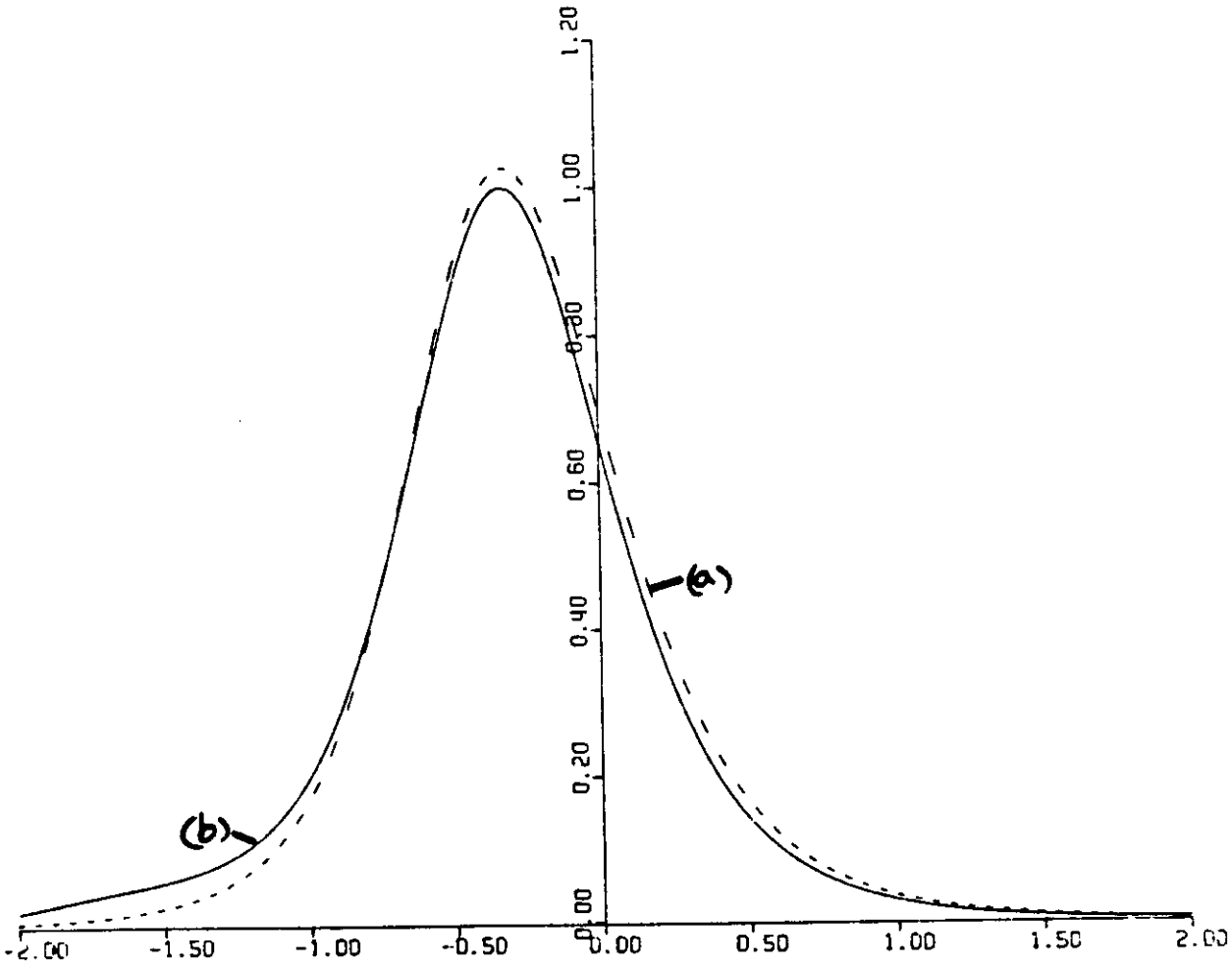

FIGURE 7 Modified Pade approximant to pdf(x): first change of coefficients

(a) exact density

(b) modified Padé approximant

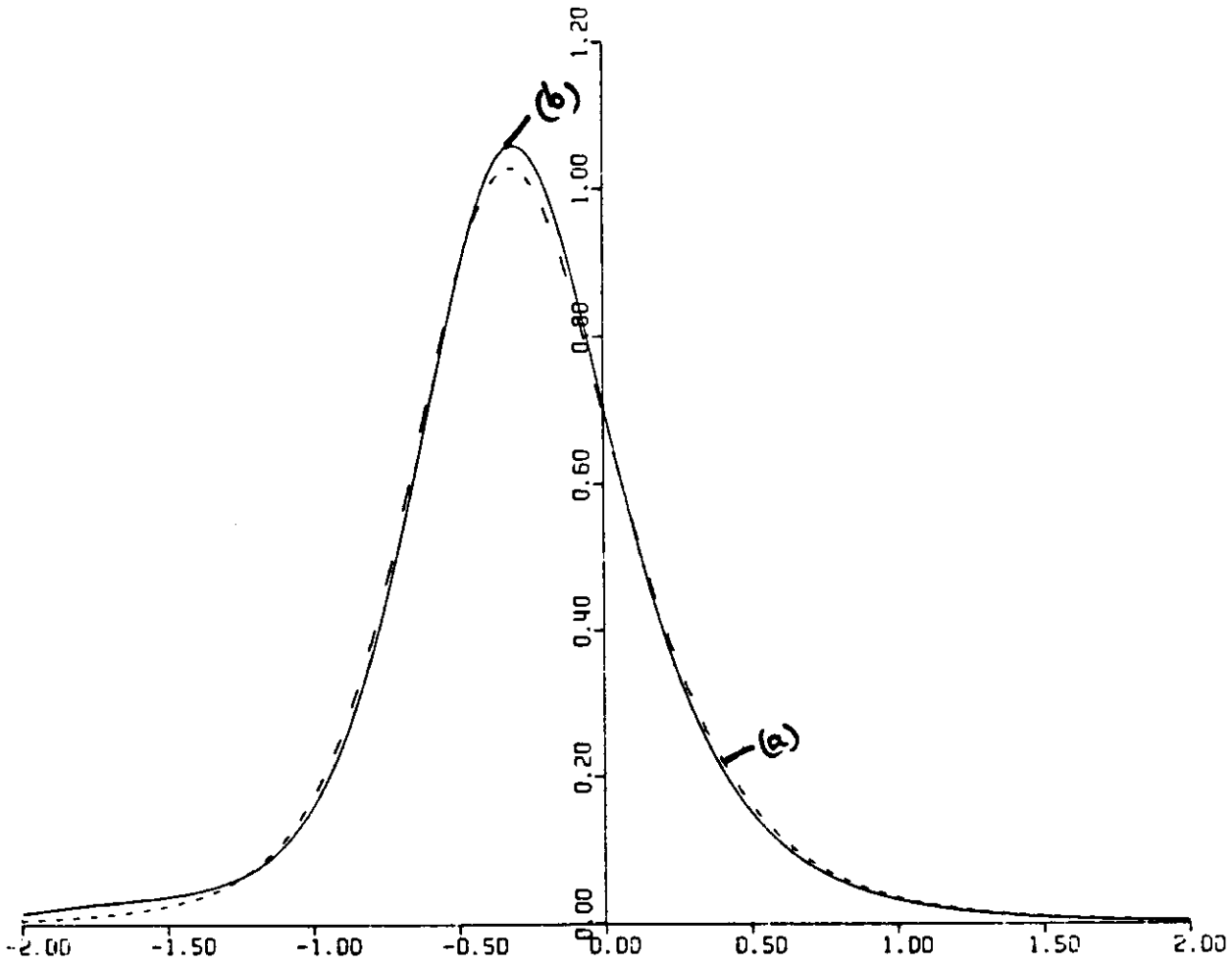

FIGURE 8 Modified Pade approximant to pdf(x): second change of coefficients

(a) exact density

(b) modified Padé approximant

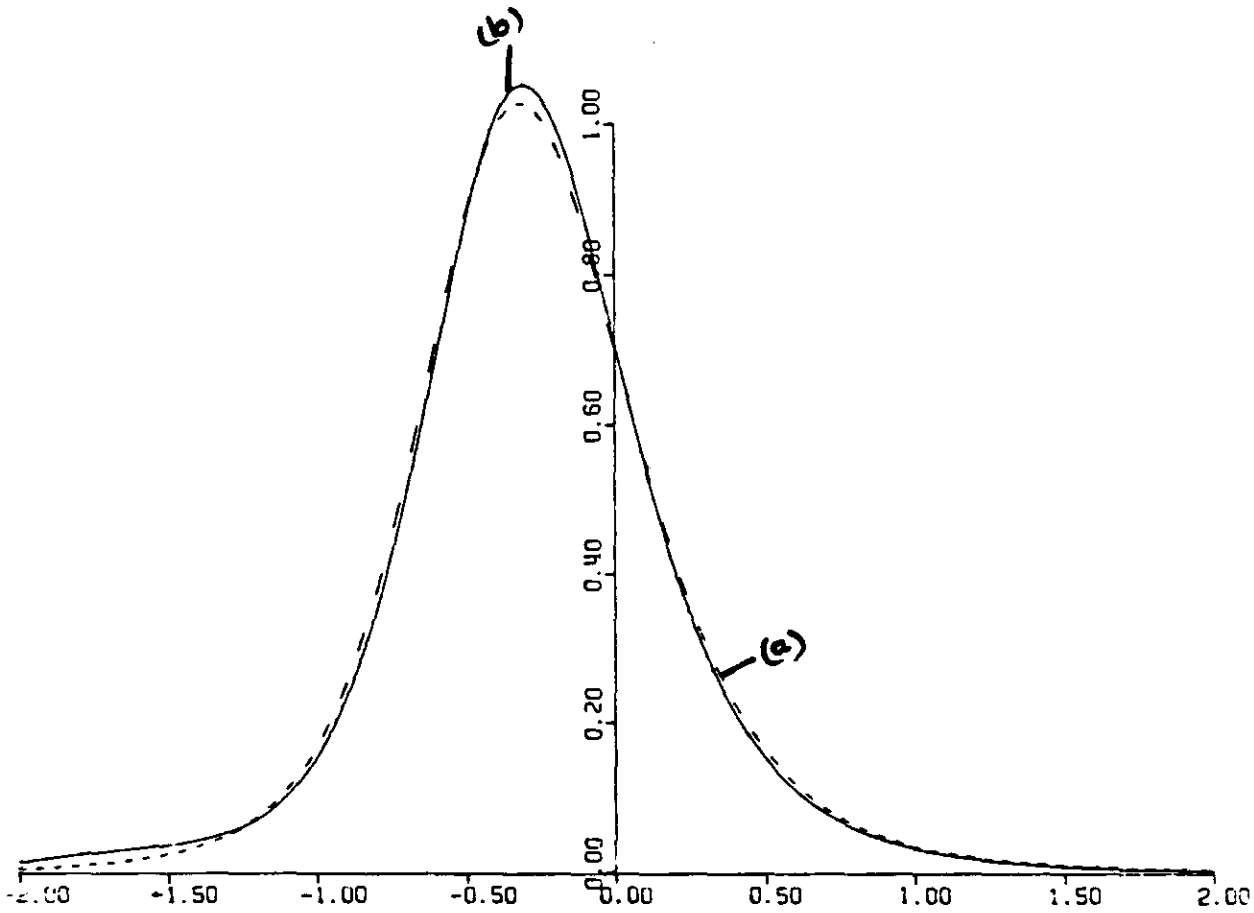

FIGURE 9 Modified Pade approximant to pdf(x) with renormalization

(a) exact density

(b) modified Padé approximant

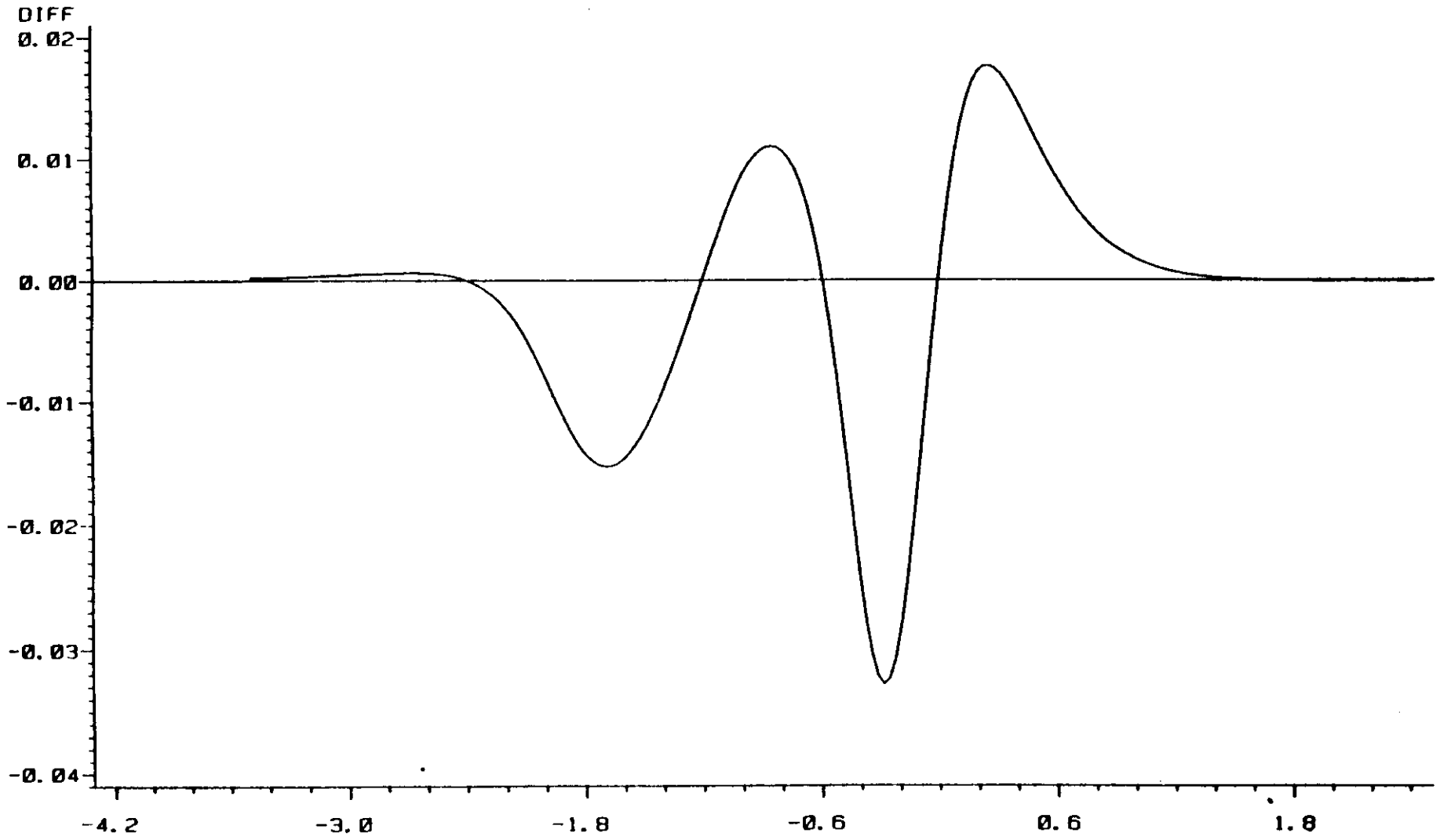

FIGURE 10: Error Curve of $R_{4,4}^{(b)}(x)$

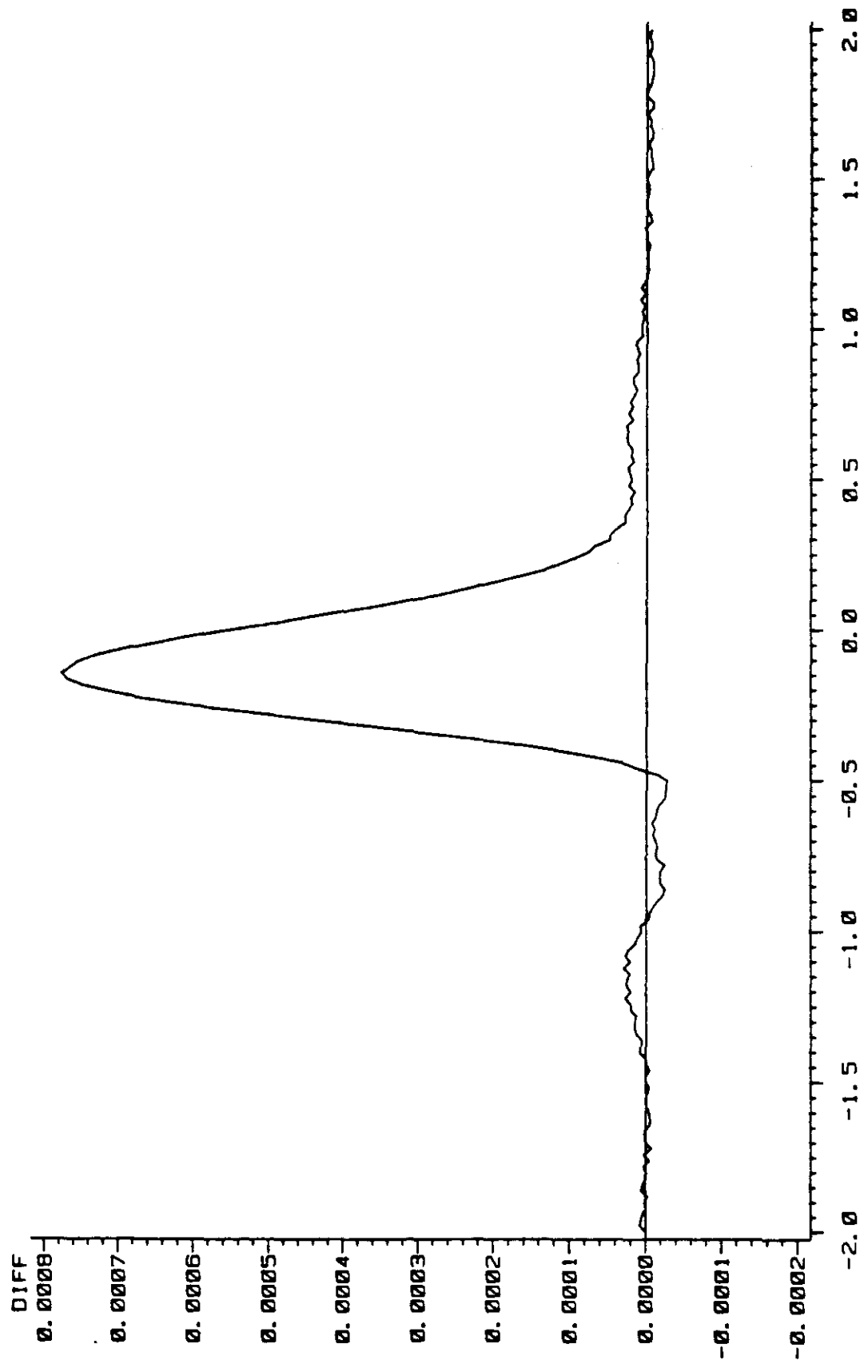

FIGURE 11: Error Curve of $R_{4,4}^{(c)}(x)$

REFERENCES

[1] Achieser, N. I., "Theory of Approximation," New York: Frederick Ungar, 1956.

[2] Anderson, T. W. and T. Sawa, "Distribution of Estimates of Coefficients of a Single Equatin in a Simultaneous System and Their Asymptotic Expansions," Econometrica, 41, 1973, pp. 683-714.

[3] _____, "Evaluation of the Distribution Function of the Two-Stage Least Squares Estimate," Econometrica, 47.

[4] Anderson, T. W., K. Morimune, and T. Sawa, "The Numerical Values of Some Key Parameters in Econometric Models," Stanford University Technical Report No. 270 (IMISS).

[5] Baker, G. A., The Essentials of Padé Approximants, Academic Press, 1975.

[6] Baker, G. A. and J. L. Gammel, The Padé Approximant in Theoretical Physics, Academic Press, 1970.

[7] Barndorff-Nielson, O. and D. R. Cox, "Edgeworth and Saddlepoint Approximations with Statistical Applications," Journal of Royal Statistical Society, 1979.

[8] Basmann, R. L., Exact Finite Sample Distributions and Test Statistics: A Survey and Appraisal," Chapter 4 in M. D. Intriligator and D. A. Kendrick, Frontiers of Quantative Economics, Vol. II, Amsterdam: North-Holland, 1974.

[9] Chui, C. K., "Recent Results on Padé Approximants and Related Problems," In G. G. Lorenz, et al., Approximation Theory II, Academic Press, 1976.

[10] Daniels, H. E., "Saddlepoint Approximations in Statistics," Annals of Mathematical Statistics, 25, 1954, pp. 631-650.

[11] _____, "The Approximate Distribution of Serial Correlation Coefficients," Biometrika, 43, 1956, pp. 169-185.

[12] _____, "Exact Saddlepoint Approximations," Biometrika, 1980.

[13] Durbin, J., "The Approximate Distribution of Partial Serial Correlation Coefficients Calculated from Residuals from Regression on Fourier Series," Biometrika, 1980.

[14] _____, "Approximations for Densities of Sufficient Statistics," Biometrika, 1980.

[15] Erdéyli, A., Asymptotic Expansions, Dover, 1956.

[16] Feller, W., An Introduction to Probability Theory and Its Applications, Vol. II, Wiley, 1971.

[17] Freud, G., D. J. Newman and A. R. Reddy, "Chebyshev Rational Approxito $e^{-|x|}$ on the Whole Real Line," Quarterly Journal of Mathematics Oxford, 1977.

[18] Frobenius, G., "Ueber Relationem Zwischen den Näherungsbruchen von Potenzreihen," Journal für die Reine und Angewandte Matematik, 90, 1881, pp. 1-17.

[19] Hart, J. F., Computer Approximations, Wiley, 1968.

[20] Hastings, C., Approximations for Digital Computers, Princeton, 1955.

[21] Holly, A., and P. C. B. Phillips, "A Saddlepoint Approximation to the Distribution of the k-Class Estimator in a Coefficient in a Simultaneous System," Econometrica, 47, 1979, pp. 1527-1548.

[22] Kendall, M. G. and A. Stuart, The Advanced Theory of Statistics, Vol. I, London: Griffin, 1969.

[23] Lighthill, M. J., An Introduction to Fourier Analysis and Generalized Functions, Cambridge University Press, 1958.

[24] Lukacs, E., Characteristic Functions, London: Griffin, 1970.

[25] Meinardus, G., Approximation of Functions: Theory and Numerical Methods, Vol. 13, Springer, 1967.

[26] Newman, D. J., "Rational Approximation to $|x|$," Michigan Mathematical Journal, 11, 1964, 11-14.

[27] Oberhettinger, F., Fourier Transforms of Distributions and Their Inverses, Academic Press, 1974.

[28] Padé, H., "Sur la Representation Approchee d'unefonction par des Fractions Rationelles," Annales Scientifiques de l'Ecole Normale Superievre, 9, 1892, pp. 1-93 (supplement).

[29] Phillips, P. C. B., "Approximations to Some Finite Sample Distributions Associated with a First Order Stochastic Difference Equation," Econometrica, 45, 1977a, pp. 463-486.

[30] _____, "A General Theorem in the Theory of Asymptotic Expansions as Approximations to the Finite Sample Distributions of Econometric Estimators," Econmetrica, 45, 1977b, pp. 1517-1534.

[31] Phillips, P. C. B., "Edgeworth and Saddlepoint Approximations in the First-Order Non-Circular Autoregression," Biometrika, 65, 1978, pp. 91-98.

[32] _____, "A Note on the Saddlepoint Approximation in a First Order Noncircular Autoregression," Cowles Foundation Discussion Paper No. 487, Yale University, 1978.

[33] _____, "Finite Sample Theory and the Distributions of Alternative Estimators of the Marginal Propensity to Consume," Review of Economic Studies, 47, 1980a, pp. 183-224.

[34] _____, "The Exact Finite Sample Density of Instrumental Variable Estimators in an Equation with n+1 Endogenous Variables," Econometrica, 48, 1980b, pp. 861-868.

[35] _____, "Best Uniform Approximation to Probability Densities in Econometrics," Invited paper presented to the World Congress of the Econometric Society at Aix-en-Provence, 1980.

[36] _____, "Chebyshev Approximation by Rational Functions on $(-\infty, \infty)$," In preparation.

[37] Rice, J. R., The Approximation of Functions, Volumes I and II, Addison Wesley, 1964 and 1967.

[38] Richardson, D. H., "The Exact Distribution of a Structural Coefficient Estimator," Journal of the American Statistical Association, 63, 1968, pp. 1214-1226

[39] Rudin, W., Real and Complex Analysis, McGraw Hill, 1974.

[40] Runge, C., "Zur theorie der ein deutigen analytischen funktioner," Acta Mathematics, 6, 1885, pp. 229-244.

[41] Saff, E. B. and R. S. Varga, Padé and Rational Approximation, Academic Press, 1977.

[42] Sargan, J. D., "Econometric Estimators and the Edgeworth Approximation," Econometrica, 44, 1976, pp. 421-428; "Erratum," Econometrica, 15, p. 272.

[43] _____, "On Monte Carlo Estimates of Moments Which are Infinite," Advances in Econometrics, (forthcoming 1980).

[44] Sargan, J. D., and W. M. Mikhail, "A General Approximation to the Distribution of Instrumental Variable Estimators," Econometrica, 39, 1971, pp. 131-169.

[45] Sargan, J. D., and Y. K. Tse, "Some Experience of Numerical Computation of Edgeworth Approximations," Mimeographed LSE (1980)

[46] Slater, L. J., Confluent Hypergeometric Functions, Cambridge University Press, 1960.

[47] Tchebycheff, P. L., "Sur les questions de minima qui se rattachent a la représentation par la methode des moindres carrés," Oeuvres I, pp. 273-378.

[48] Timan, A. F., Theory of Approximation of Functions of a Real Variable. New York: MacMillan (1963).

[49] Walsh, J. L., "The Existence of Rational Functions of Best Approximation," Transactions of the American Mathematical Society, 33, 1931, 668-689.

[50] _____, "Interpolation and Approximation by Rational Functions in the Complex Domain," American Mathematical Society, 1965 (originally published 1935).

[51] Zelen, M., and N. C. Severo, "Probability Functions," Chapter 26 in M. Abramowitz and I. A. Stegun, Handbook of Mathematical Functions, Dover, 1965.